Assessing the Utility of DMSP-OLS night-time images to propose surrogate census

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Koel Roychowdhury

MSc Geography (Hons)

MSc GIS & Human Geography

School of Mathematical and Geospatial Sciences

College of Science, Engineering and Health

RMIT University

August 2011

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Signature: …………………………

Name: Koel Roychowdhury

Date: …………………………

Acknowledgements

The research was conducted under the Australian Federal Government funded Australian Leadership Award Scholarship (ALAS) Scheme. This thesis would not have been possible without the contribution of so many people in many different ways.

Firstly, my special thanks go to my supervisors Prof. Simon Jones and Dr. Colin Arrowsmith and my advisor Dr. Karin Reinke. I gratefully acknowledge all their support and advices over the course of my PhD. I consider myself to be extremely fortunate to get them as my supervisors. Thank you to Simon for introducing me to the night-time datasets. I am also thankful to Simon and Karin for all their detailed comments and feedbacks which significantly helped to improve the drafts. I could not have completed this thesis without your words of encouragement and invaluable guidance.

I would like to thank Dr. Benjamin Tuttle from NGDC, NOAA for providing me the required datasets and clearing my doubts on them. My special thanks go to Dr. Navalgund, Director, Space Application Centre, India and Dr. S.K. Pathan, Project Director, NRDB and Head, GIDD, Space Application Centre, India for giving me the privilege of using the spatial datasets of the administrative regions for this study. I would also like to thank Dr. Anthony Bedford for going through all the statistical methods used in this research.

Thanks to all my fellow researchers in the "Yellow Lab" in RMIT; especially to Danni, Kat, Alex, Naoko, Grant, Rakhesh, Liz, Steve, Samia and Vaibhav for providing such a supportive and fun –filled ambience to share ideas and experiences. You all made the "Yellow Lab" my second home for the last three years. Vaibhav, thank to you very much for all your support in one of the most stressful times. My special thanks to Liz for going through some important parts of this thesis.

I wish to thank all my friends for their contribution. My special thanks to Vaishnav and Taz , Pooja and Chaitali for all those moments of laughter. Marie, thank you for the coffee breaks in Starbucks. My earnest gratitude to Towfiq, NeelDa, Mou and AyonDa, for being there and helping me in so many different ways - from giving a patient hearing to uplifting my mood. I couldn't have managed without you.

Finally to my parents, a "Thank You" seems so very inadequate but without you I could have never reached to this stage. To Ma (Mum) and Baba (Dad), thank you for your love, encouragement, support, trust and confidence. I'm coming back to you Ma!

Table of Contents

List of Figures

List of Tables

List of Equations

List of Acronyms

AFFRIT	Agriculture, Forestry and Fisheries Information Technology Centre
AFWA	Air Force Weather Agency
ASGC	Along Scan Gain Control
AVHRR	Advanced Very High Resolution Radiometer
AVIRIS	Airborne Visible/Infrared Imagining Spectrometer
AWiFS	Advanced Wide Field Sensor
BRDF	Bi-Directional Reflectance Distribution Function
CD	Community Development
CDIAC	Carbon Dioxide Information Analysis Centre
CO_2	Carbon Dioxide
CTM	Chemical Transport Model
DDP	District Domestic Product
DMSP	Defense Meteorological Satellite Programme
DN	Digital Number
DNB	Day/Night Band
DoD	Department of Defense
EDES	Early Damage Area Estimation System
EIFOV	Effective Instantaneous Field of View
EMS	Electro Magnetic Spectrum
EOG	Earth Observation Group
ETM	Enhanced Thematic Mapper
FAO	Food and Agricultural Organization
FOV	Field of View
FWHM	Full Width at Half Maxima
GCPs	Ground Control Points
GDP-PPP	Gross Domestic Product- Purchasing Power Parity
GIS	Geographic Information System
GOME	Global Ozone Monitoring Experiments
GRP	Gross Regional Product
GRUMP	Global Rural Urban Mapping Project
GSD	Ground Sampling Distance
ICR	Intelligent Character Recognition
IDI	Information and Communication Technology Development Index
IFOV	Instantaneous Field of View
IRS	Indian Remote Sensing Satellite

ISA	Impervious Surface Area
ISRO	Indian Space Research Organization
ISS	International Space Station
LEDs	Light-Emitting Diodes
LIDAR	Light Detection and Ranging
MAUP	Modifiable Areal Unit Problem
MMA	Mumbai Metropolitan Area
MODIS	Moderate Resolution Imaging Spectroradiometer
MSS	Multi Spectral Scanner
NASA	National Aeronautics and Space Agency
NCEP	National Centre for Environmental Prediction
NDVI	Normalized Difference Vegetation Index
NGDC	National Geophysical Data Centre
NIC	National Informatics Centre
NOAA	National Oceanic and Atmospheric Administration
NO_x	Nitrogen Oxides
N-P	Normal Probability
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NSIDC	National Snow and Ice Data Centre
OLS	Operational Linescan System
ORGI	Office of the Registrar General, India
PCDDP	Per Capita District Domestic Product
PMT	Photo Multiplier Tube
PSF	Points Spread Function
RCC	Regional Computer Centre
RMSE	Root Mean Square Error
SAC	Space Application Centre
SC	Schedule Cast
SEP	Subtotal Ecological-Economic Product
SIDaB	Satellite Image Database System
SMC	Space and Missile System Centre
SPIDR	Space Physics Interactive Data Resource
SPOT	Systeme Probatoire d'Observation de laTerre
SPOT VGT	Systeme Probatoire d'Observation de laTerre Vegetation
SRTT	Serial Rolling Total Tabulator cum Printer
SS	Sum of Squares
ST	Schedule Tribe

SVM	Support Vector Machine
TIN	Triangulated Irregular Network
TM	Thematic Mapper
UAs	Urban Agglomerations
UN	United Nations
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VDGA	Variable Digital Gain Amplifier
VIIRS	Visible/Infrared Imager/Radiometer Suite
VNIR	Visible Near Infrared
WGS	World Geographic System

List of Publications:

Report:

Buxton, M, Alvarez, A, Butt, A, Farrell, S, O'Neill, D, Lechner, A & Roychowdhury, K 2008, Planning Sustainable Futures for Melbourne's Peri-urban Region, Melbourne.

Conference Proceedings (non peer-reviewed):

Roychowdhury, K., Jones, S. D. & Arrowsmith, C. 2009. Assessing the utility of DMSP/OLS night-time images for characterizing Indian Urbanization. Joint Urban Remote Sensing Event. Shanghai, China.

Journal article (peer- reviewed):

Roychowdhury, K., Jones, S. D., Arrowsmith, C. & Reinke, K. 2011. A comparison of high and low gain DMSP/OLS satellite images for the study of socio-economic metrics. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4, 35-42.

Conference Proceedings (peer-reviewed):

Roychowdhury, K., Jones, S. D. & Arrowsmith, C. 2009. Mapping Urban Areas of India from DMSP/OLS Night-time Images. In: Ostendorf, B., Baldock, P., Burdett, M. & Corcoran, P., eds. Surveying & Spatial Sciences Institute Biennial International Conference, 2009 Adelaide, Australia. Surveying & Spatial Sciences Institute, 319 - 331.

Roychowdhury, K., Jones, S. D., Arrowsmith, C., Reinke, K. & Bedford, A. 2010. Estimating census metrics at a sub-national level using radiance calibrated DMSP-OLS night-time images. In: Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Roychowdhury, K., Taubenboeck, H. & Jones, S. D. 2011. Delineating urban, suburban and rural areas using Landsat and DMSP-OLS night-time images. In: Urban Remote Sensing Event (JURSE), 2011 Joint, 11-13 April 2011. 33-36.

Roychowdhury, K., Jones, S. D., Arrowsmith, C. & Reinke, K. 2011. Indian census using satellite images: Can DMSP-OLS data be used for small administrative regions? In: Urban Remote Sensing Event (JURSE), 2011 Joint, 11-13 April 2011. 153-156.

Bhandari, L. & Roychowdhury, K. 2011. Night Lights and Economic Activity in India: A study using DMSP-OLS night time images. Asia-Pacific advanced Network (APAN) Workshop. New Delhi, India.

Abstract:

The collection of population census is an essential component for the monitoring and management of resources and hence the health and prosperity of its residents. This research proposes a novel surrogate measure for the collection of key census attributes using a case study from western India. Satellite images from the Operational Linescan System (OLS) onboard the Defense Meteorological Satellite Program (DMSP) group of satellites were used for the study. Unlike other passive remote sensing sensors, this sensor is capable of recording the emissions from artificial lights on the earth surface. Radiance calibrated images and stable light products from DMSP - OLS satellite captured in 2001 are used in this study. This year corresponds with the year of the last available full Indian census. The research focuses on the state of Maharashtra, the second most populated state of India. The study considers three major issues:

1. Which census metrics can be correlated with light information obtained from DMSP-OLS satellite images?
2. What is the most appropriate spatial resolution or mapping unit for attributing the census metrics?
3. Which night-time satellite image product obtained from the DMSP-OLS satellite is most suitable for the purpose of mapping surrogate census metrics?

Countries, such as India, conduct a census collection every ten years. Currently the census in India is carried out manually with enumerators visiting every household in the country. Being such a vast country (in terms of area) and with a population of more than 1 billion, manual data collection is a laborious and expensive process. The census suffers from a number of shortcomings including inconsistency issues, the Modifiable Areal Unit Problem (MAUP) and large temporal acquisition timeframes. Although, provisional population figures are published in the same year as census enumeration, the final tables require more than a year of processing prior to publication. This thesis proposes a surrogate method for collecting key census metrics based on satellite images. Such surrogates overcome some of the issues previously identified and enable the prediction of census metrics more frequently than is currently available. This in turn can provide valuable information to support population policy making and development planning.

In order to determine which socio-economic census metrics were correlated with light information obtained from DMSP-OLS images, the analysis initially considered 144 variables. These included all the demographic metrics and some socio-economic variables such as amenities, education facilities and power source. Initial statistical analyses (for example, histogram, skewness – kurtosis and

goodness of fit test) examined the normality of distribution. From these tests 48 out of 144 metrics were confirmed to be normally distributed. These were then used for bootstrapping the correlation coefficients. Correlation coefficients between these 48 variables and metrics extracted from the DMSP-OLS images were calculated from 1000 bootstrap samples at the 95% confidence interval. Variables with a bias of less than 0.05 and standard error of less than 0.2 were selected for further analyses. Ten census metrics met these criteria; these included some common demographic variables such as population density, *Per Capita District Domestic Product* (*PCDDP*) and percentage of households with access to cars, jeeps and vans.

A few data quality issues such as those specifically related to positional accuracy, attribute accuracy, absence of metadata and data lineage were encountered in this research. These were likely due to relative positional inconsistency between DMSP-OLS images and the vector datasets of administrative regions in addition to attribute errors associated with the taluks and village datasets. Solutions such as geometric corrections were applied during the analyses to manage the data quality issues identified.

To determine the most appropriate spatial resolution or mapping unit for attributing the census metrics (question 2), the significant r values from correlations between chosen metrics and the mean and standard deviation of radiance, stable lights and brightness information as obtained from satellite images were compared. These analyses were conducted independently at three spatial scales or administrative regions as defined by the Indian census (districts, taluks or sub–districts and villages). At the district level, non–composited fixed gain radiance calibrated images were used along with the global composite images. Two fixed gain images were used: (a) a low gain image with a gain setting of 20 dB and (b) a high gain image with a gain setting of 50 dB. The ten census metrics were demonstrated to be correlated with mean and standard deviation of radiance, brightness and stable lights. The ‘r’ values were significant ranging from 0.6 to 0.9 at the 95% confidence interval. Calculations for taluks and villages used only the global composite images. At the taluk level, significant ‘r’ values ranged from 0.4 to 0.6 at the 95% confidence interval. No significant correlations were found at village level. From the correlation coefficients it was concluded that DMSP-OLS night-time images were most suitable for proposing census at the district level (areas ranging from 150 Km^2 to more than 15,000 Km^2).

Multiple images are available from the DMSP-OLS satellite including stable light image, average digital number image and radiance calibrated image. Determining the most suitable image for the purpose of proposing surrogate census, models were proposed and adjusted r^2 compared independently for all regional scales. Linear regression and multivariate analyses were performed and models proposed for each of the selected census metrics. Results ranged from r^2 of 0.8 to 0.9 at the 95% confidence interval at the district level. At taluks the adjusted r^2 values ranged from 0.2 to 0.8 at 95%

confidence interval, with the majority of the metrics being moderately correlated (with r^2 between 0.4 and 0.7). Generally it was found that the observed lights and brightness of large rural settlements from DMSP-OLS images had the potential to predict particular census metrics. However, unlike larger areas such as districts where DMSP-OLS night-time images adequately predicted census metrics, at the sub-district level the results needed to be supplemented and validated with ancillary information sources such as survey reports. No significant correlations were identified at the village level and this was attributed to sampling issues. All models were validated over eight withheld districts and 90 withheld taluks. Optimum models which proposed the metrics within a 25% error margin were selected. For those metrics with more than one optimum model, the one with the highest adjusted r^2 value was chosen. Residual maps were then used to assess the models. It was demonstrated that all the models were more than 60% accurate with some models, such as those for *PCDDP* and *number of households per square kilometre*, achieving an accuracy of greater than 85%. Metrics currently unavailable in traditional census at spatial scales lower than the district level were also predicted using the proposed models. Maps showing the predicted measures were also generated.

This thesis proposes a surrogate method for the collection of key census metrics using satellite images. Such surrogates aim to overcome some of the problems and limitations imposed by manual census collection in developing countries. The method proposed in this study will enable the rapid prediction of census metrics more frequently than is currently available. This will support population policy making, the updating of resource inventories and development planning. A further benefit of this method is the potential to predict census metrics at spatial scales currently unavailable. This includes taluks, villages and areas as small as approximately 1 square kilometre (or the area represented by one pixel in the DMSP-OLS satellite images). The models proposed in this research may also be applied to predict census metrics within user defined areas such as urban agglomerations. The results obtained from this research demonstrate the value and utility of DMSP-OLS night-for estimating key census metrics in near-time and across multiple scales.

1. Introduction

The research presented in this thesis looks into the problems of conducting manual census and proposes a method for surrogate census using remote sensing techniques. The census is an essential component for the monitoring and management of population and resources. Satellite images from the Operational Linescan System (OLS) onboard the Defense Meteorological Satellite Program (DMSP) group of satellites were used in the study. This chapter provides an overview of the benefits of a census, problems of conducting a manual census over a vast country such as India and outlines the benefits of surrogate census using remote sensing techniques. The research questions and overall structure of this thesis will also be presented.

1.1: Background:

A population census is described as a complete enumeration of people which is conducted every five or ten years (Martin, 2006). In addition to being a headcount of the people, a census is an inventory of information on demography, economic activities, literacy and education, migration, housing and household amenities as well as social composition of the population such as religion, tribes and language (Office of the Registrar General and Census Commissioner, 2011b). A census has the advantage of comprehensive population coverage over fine geographical details (Martin, 2006). Census data is used in a variety of user applications such as policy implementation, resource allocation and administrative processes. In India census data is used in delimitation of Parliamentary, Assembly and local body Constituencies (Office of the Registrar General and Census Commissioner, 2011c). However, a traditional census suffers from a large number of problems such as enumeration difficulties, low response rates, disclosure control, large operational expenses, organizing enumerators and distribution and collection of census questionnaires to and from every household (Martin, 2006, Martin, 2000). The collection and reporting geographical units are the same for a traditional census. This makes the application of the data difficult for areas smaller than these units or for administrative areas that do not share the same boundaries (Martin, 2006, Duke-Williams and Rees, 1998).

India is the second most populated country in the world with a population of around 1.2 billion in 2011. The Indian census is conducted once every ten years. It is one of the largest population counting activities in the world which takes several months to complete. The enumeration for the latest 2011 census in India engaged 2.7 million enumerators who individually visited 240 million households all over the country. There were 340 million questionnaires in 18 different languages specially designed and printed with bar coding and unique form numbering. These were accompanied by 540 million instruction and training materials. The questionnaires and instruction manuals were distributed to 17,000 distinct locations all over the country for conducting the enumeration. Uniform training was provided to all the 2.7 million enumerators by 54,000 trainers. For the purpose of digitally processing the data, 16 workstations were set up all over the country with skilled operators and the latest hardware and software facilities (Office of the Registrar General and Census Commissioner, 2011b). The census was conducted in two phases, the first phase was the house-listing and housing census and the second phase was the population enumeration. The total cost involved in the census was $490 million (BBC News, 2011). In addition to being such an intensive process, both administratively and economically, the Indian census suffers from a number of shortcomings including inconsistency issues, the Modifiable Areal Unit Problem (MAUP) and large temporal acquisition and publication timeframes. Although, the provisional population figures are published in the same year of census enumeration, the final tables require more than a year to be made available to the people. The provisional tables are presented by adding up the population data collected by each census enumerator. These tables suffer from errors in calculation and duplication or omission of enumeration units (Office of the Registrar General and Census Commissioner, 2011a). There is also no database on the census metrics for the inter-censal period. Furthermore, for smaller administrative regions, some census metrics are not collected.

In order to overcome the problems of a manual census, this research proposes a surrogate census at subnational administrative regions. DMSP-OLS night-time images were used in this study. Figure 1.1 shows a DMSP-OLS image of the whole world.

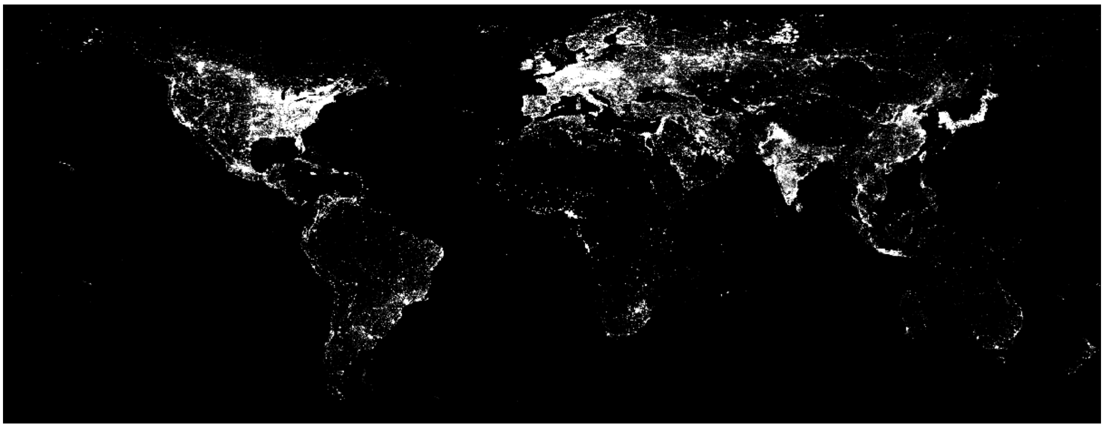


Figure 1.1: Example of DMSP-OLS night-time image of the world

A look at the night-time lights imagery (figure 1.1) shows how conspicuous the relationship between population distribution and night-time lights is – the brightly lit areas are the more inhabited parts of the world (for example, the United States, Western European countries, India and Japan), and the areas with deficiency in lighting are the less populated areas of the world (e.g., Amazon Basin and Central

Australian desert). DMSP-OLS imagery has been widely used to study population density, urbanization, green house gas, fire and economic development (Doll et al., 2002, Elvidge et al., 2003, Elvidge et al., 2004, Elvidge et al., 2005, Elvidge et al., 2007a, Imhoff et al., 1997b, Sutton et al., 1997, Sutton et al., 2001, Chand et al., 2009). The variation in the intensity of light as recorded from the DMSP-OLS images of India were used in this research to propose models to predict census metrics for small geographical areas. The method proposed in this study will enable the prediction of census metrics more frequently than available now. This in turn will provide more timely and frequent information needed for population policy making and development planning. Another benefit of the proposed method is that it will potentially enable the prediction of unavailable census metrics at the sub-district level as well as areas as small as approximately one square kilometre (i.e. the area represented by one pixel in the satellite images used).

1.2: Research Questions:

This research thesis will develop a method that will supplement traditional census counts in developing countries. In order to propose a surrogate method for collecting key census metrics from DMSP-OLS images, three research questions were investigated. These are as follows:

RQ1: Which socio-economic metrics obtained from the census can be correlated with light information obtained from DMSP-OLS satellite images?

RQ2: What is the most appropriate spatial resolution or mapping unit for attributing the census metrics using DMSP-OLS night-time images?

RQ3: Which night-time satellite image product obtained from the DMSP-OLS satellite is most suitable for the purpose of mapping surrogate census metrics?

1.3: Thesis Structure and outline of chapters

The section presents the structure of the thesis in figure 1.2.

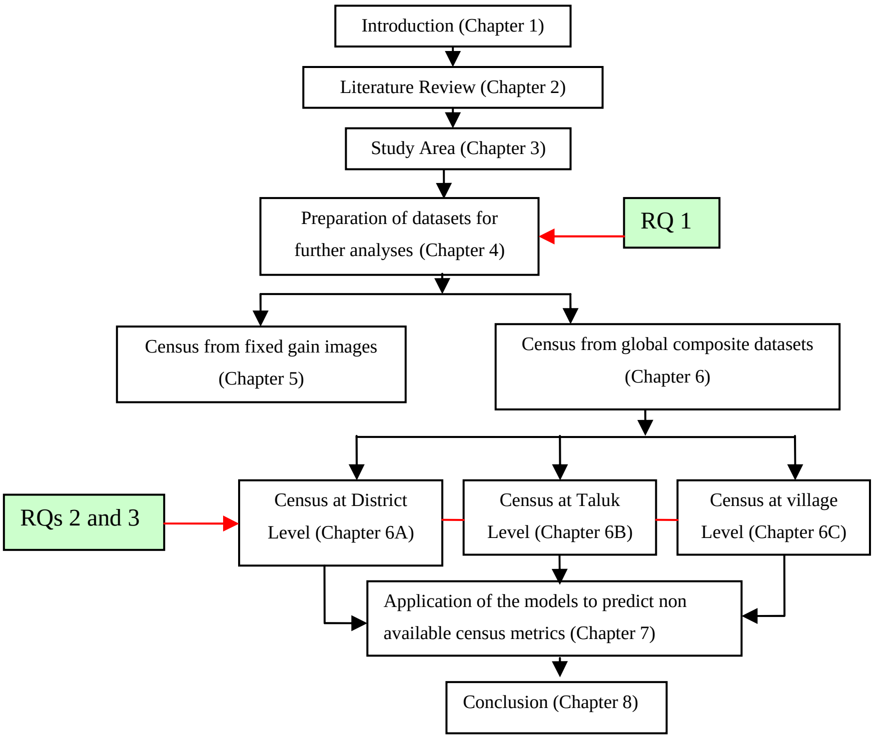


Figure 1.2: Structure of the thesis

1.3.1: Outline of Chapters:

The thesis is presented in eight chapters.

1. Introduction

The first chapter introduces the research and provides a background and rationale for the study. The chapter explains the research aim and objectives. It puts forward the research questions and the structure of the thesis in the form of outline of the different chapters.

2. Literature Review

Chapter 2 provides a background to the research by critical review of the literature on satellite remote sensing and census. Firstly it describes the characteristics of different DMSP-OLS night-time images. Following this, this chapter provides an insight into the Indian census and its drawbacks. The chapter further reviews the application of satellite remote sensing in predicting census metrics at different spatial scales with special reference to various applications of DMSP-OLS images.

3. Study Area

Chapter 3 gives an overview of the study area for this research. The study looked into the state of Maharashtra in western India. The state exhibits diversity in urbanization and distribution of population. A brief idea on the state as obtained from the census is presented in this chapter.

4. Preparation of datasets for further analyses

Chapter 4 describes the pre-processing of the census and satellite datasets used in this research and the various data quality issues that were encountered. There are four parts in this chapter. The first part describes census data processing and the different statistical tests conducted on the available census metrics. The second part describes the satellite image processing where image pre – processing and image analyses steps are discussed in detail. In the third part of the chapter, the details of the data quality elements are addressed. The final part of the chapter combines the results from the first two parts. This chapter answers research question 1 and shortlists the census metrics suitable for further analyses.

5. Scope and Limitations of creating a surrogate census from fixed gain radiance calibrated images

This chapter describes the usefulness of fixed gain radiance calibrated DMSP-OLS night-time images in proposing a surrogate census. Firstly, the chapter analyzes the correlations between radiance and census metrics. Several predictive models are subsequently proposed. These are further examined in detail and validated over the withheld districts. The chapter concludes with the selection of optimum gain and the limitations of single orbit fixed gain images.

6. Surrogate census from global composite images at district, taluk and village level

This chapter uses the stable light and brightness data from DMSP-OLS satellite to propose surrogate census for areas at different spatial scales. The chapter is divided into three parts. Part 6A describes the results at the district level. Part 6B deals with results at the taluk level and the results at the village level are presented in part 6C. After examining the correlations of the census metrics with mean and standard deviation of brightness and stable lights at districts, taluks and villages, the chapter examines linear regression models and multiple regression models. The chapter concludes with optimum models for each spatial scale. This chapter answers research questions 2 and 3.

7. Applications of the models to produce maps of non-available census metrics for small regions

Chapter 7 applies the models proposed in chapter 6 to produce the maps of non-available census metrics for small regions. Maps are produced for villages and for areas as small as one square kilometre. Firstly the chapter describes errors associated with Modifiable Areal Unit Problem (MAUP) and ecological fallacy followed by the maps for the villages and at one square kilometre area. The chapter concludes with an overall assessment of the results at various spatial scales.

8. Conclusion

Chapter 8 draws together the overall results. The thesis concludes with the limitations of this research and suggestions for future work.

2. Literature Review

In order to explore the application of remote sensing, especially night-time images in socio-economic and population studies, this chapter examines the nature of the DMSP-OLS satellite and the different products obtained from it (section 2.1). Following a description of Indian census (section 2.3), this chapter describes the application of remotely sensed techniques and night-time images in census and other socio – demographic applications (sections 2.4 and 2.5).

2.1: The Defense Meteorological Satellite Programme – Operational Linescan System (DMSP-OLS)

2.1.1: The Satellite:

The Defense Meteorological Satellite Programme (DMSP) is a Department of Defense (DoD) program of US Air Force Space and Missile Systems Centre (SMC) (National Geophysical Data Centre, 2007a). The mission was launched in mid 1960s (High Energy Astrophysics Archive Research Centre, 2003). The meteorological, oceanographic and solar-terrestrial physics environments monitoring satellites are designed, built, launched and maintained by this program from an altitude of approximately 800 Km (National Geophysical Data Centre, 2007a). The archive of the DMSP-OLS dataset is maintained by the Earth Observation Group (EOG) at the National Geophysical Data Centre at National Oceanic and Atmospheric Administration (NGDC/NOAA) (National Geophysical Data Centre, 2007a). DMSP-OLS data are downloaded by the Thule Air Force Base from the satellite constellation (http://www.thule.af.mil/) and transmitted to the Air Force Weather Agency (AFWA) via communication satellites. The data are decrypted at the AFWA and are sent to the NGDC via T1 carriers. In NGDC, this data are further processed using dedicated softwares. The processing involves a satellite renavigation, followed by re – ordering and splitting of the data. The data problems from transmission are fixed before they are organized into orbits and written into a library of robotic tapes (Frederik, 2002).

The DMSP has seven sensors, each used for separate applications. The different types of sensors present on a DMSP satellite and their applications are summarized in the table (table 2.1) below:

Table 2.1: DMSP sensors and the applications of their data (National Geophysical Data Centre, 2007b)

Sensors of DMSP satellites	**Applications**
OLS – Operational Linescan System: collects data in the visible and infrared portion of the electromagnetic spectrum	– Records the visible and near infrared emissions from cloud tops. – Also capable of recording thermal emissions on the earth surface (man-made and natural) and northern lights. – It also records aurora, city-lights, gas flares and man-made and natural gas flares.
SSM/I – Microwave Imager	– Derive geophysical parameters such as ocean surface wind speed, area covered by ice, age of ice, ice edge, precipitation over water, soil moisture, land surface temperature, snow cover and snow surface temperature.
SSMT/2 – Atmospheric Water Vapor Profiler	– It is capable of capturing the mid-latitude weather phenomena such as fronts and extra-tropical cyclones (along with their three-dimensional structure). – Tropical cyclones, tropical plumes, subtropical anticyclones and sea ice and snow cover may be identified from these images.
SSJ/4 – Precipitating Electron and Ion Spectrometer	– Used in space physics research, monitoring the space environment and space weather forecasting, the structure of high latitude ionosphere. – It has been used to characterize and model the polar cap, the auroral zone (the equatorward boundary) and the South Atlantic Geomagnetic Anomaly.
SSM/T – Atmospheric Temperature Profiler	– Atmospheric temperature sounding.
SSI/ES – Ion Scintillation Monitor	– NA
SSM - Magnetometer	– In combination with SSI/ES and SSJ/4, it provides heating and electron density profiles in the high latitude ionosphere.

Of these sensors, the Operational Linescan System (OLS) is sensible to the visible and infrared portion of the Electro Magnetic Spectrum (EMS). The data from this sensor are used to monitor the global distribution of clouds and cloud top temperatures each day. The archive data of DMSP-OLS consists of low resolution (2.7 Km) imagery of the globe and high resolution (0.55 Km) regional images. The low resolution pixel values are the mean of 25 high resolution values (High Energy Astrophysics Archive Research Centre, 2003). An overview of the OLS sensor may be stated as follows:

- Visible pixels have relative values ranging from 0-63. They do not have the absolute values in watts per square metre.
- OLS sensor has two telescopes and a Photo Multiplier Tube (PMT) which is used at night.
- Visible telescope is sensitive to 0.4 – 1.10 um (0.58 – 0.91 um Full Width at Half Maxima (FWHM)) and radiance values from 10^{-3} to 10^{-5} watts/cm^2/steradian.
- Infrared telescope is sensitive to radiation from 10.0 – 13.4 um (10.3 – 12.9 um FWHM).
- The DMSP satellites are in a sun-synchronous, high altitude (800 Km) of low polar orbit.
- The detectors move to and fro in a whisk-broom or pendulum type motion.
- The continuous analog signals are sampled so that the centres of the pixels geolocated on the earth surface are approximately equidistant (0.5 Km apart).
- 7325 pixels are digitized across the 1080 swath from limb to limb.

In NGDC the DMSP-OLS data are decompressed, reordered, restructured and de-interleaved (National Geophysical Data Centre, 2007a). The four body orbital mechanics programme is used to compute the satellite ephemeris with observed (and not predicted) orbital elements as input. Missing scan lines are correctly positioned. Quality assessments are made of each pixel and characterized for each complete scan line.

The analog images of the aurora are maintained by the NGDC since 1972. Other images of the OLS sensor are preserved by the National Snow and Ice Data Centre (NSIDC) from 1979 – 1992. From 1992, all the DMSP-OLS data are in digital format. The orbit header and scan – line data records are stored in an archive file (National Geophysical Data Centre, 2007c). A "smooth" resolution scan line consists of 1465 visible pixels, 1465 IR pixels, satellite ephemeris and astronomical parameters. A high resolution scan line contains 7325 pixels and the same supporting information.

The following table (table 2.2) shows the dates of operating OLS sensors on board the DMSP satellites.

Table 2.2: Lifespan of OLS sensors on board DMSP satellites

SENSOR	**F10**	**F11**	**F12**	**F13**	**F14**	**F15**	**F16**
OLS	4/12/92 - 2/8/95	4/12/92 - 4/22/95	9/25/94 - 7/31/02	4/24/95 - Present	4/28/97 - Present	1/24/00 - Present	11/04/03 - Present

Data from sensors F10 and F11 were stopped due to sensor problems and the processing of the archive was stopped for satellite F12.

The OLS is an oscillating scan radiometer. It was designed for imaging cloud cover over the earth surface. The OLS sensor has two bands: Visible and Thermal Infrared. The visible band is sensitive to wavelengths of 0.4um to 1.1um, thus encompassing the visible and near infra red portion of the EMS. The thermal infrared channel has a band pass from 10 – 13.4 um. The visible band has 6 bit quantization with pixel values ranging from 0 – 63. The thermal infrared channel has 8 bit quantization with values ranging from 0 – 255 (Elvidge et al., 1999). The OLS has a swath of approximately 3000 Km. This wide swath provides global coverage data four times a day: dawn, day, dusk and night (Elvidge et al., 1997b, Elvidge et al., 1999). DMSP platforms are stabilized using four gyroscopes (three axis stabilization) and platform orientation is adjusted using a star mapper, Earth Limb sensor and a solar detector (Elvidge et al., 1997b, Elvidge et al., 1999). The OLS sensor at night is capable of detecting radiances down to 10^{-9} watts/cm^2/sr/um in the visible band. This is four times more sensitive than the OLS visible band at daytime and Visible-Near Infrared (VNIR) bands of other sensors such as NOAA AVHRR (Advanced Very High Resolution Radiometer), Landsat TM (Thematic Mapper) and AVIRIS (Airborne Visible/Infrared Imaging Spectrometer). Figure 10.1 in Appendix shows that OLS is more sensitive than the other sensors.

DMSP-OLS records data in two spatial resolution modes. The “fine” mode of data with a spatial resolution of 0.56 Km is also called the full resolution data. There is an on-board averaging of five by five blocks of fine mode data to produce the “smooth” data with a resolution of 2.8 Km. Most of the data received by NGDC/NOAA is in the smooth spatial resolution mode (Elvidge et al., 1997b, Elvidge et al., 1999).

The OLS sensor has methods to restrain pixel dimension enlargements (Lieske, 1981). OLS has a “sinusoidal scan motion”. In the fine resolution data (0.56 Km) this motion helps to maintain a constant Ground Sampling Distance (GSD) at all scan angles. For fine resolution data, the GSD in along scan starts at 0.31 Km at nadir. It slowly rises to 0.55 Km at 1200 Km surface distance from nadir. Thereafter it decreases again at 0.5 Km by the end of scan. The detector image rotates as a function of the scan angle, and the shape, size and orientation of the OLS detectors was designed to take advantage of this rotation of the Effective Instantaneous Field of View (EIFOV) at off-nadir scan angles. In addition, the PMT electron aperture is magnetically switched (deflected) during the outer quarter of each scan, again reducing the size of the detector image on the ground surface (Elvidge et al., 1997b). “The EIFOV of the night-time VIS band fine resolution data starts at 2.2 Km at the nadir and expands to 4.3 Km at 766 Km out from the nadir. After the aperture is switched, the IFOV is reduced to 3.0 Km and expands to 5.4 Km at the far edges of the scan. Thus, the EIFOV is substantially larger than the GSD in both the along-track and along-scan direction" (Elvidge et al., 1997b p. 729).

2.1.2: OLS Gain Control:

Gain can be defined as the ratio of the output signal strength to the input signal in any system. In OLS, the full system contains both gains and losses, but there is an overall amplification of the incoming signal. There are analog preamplifiers and postamplifiers present in the OLS along with VDGA (Variable Digital Gain Amplifier). At night the gain of the PMT also adds to the gain of the system. The VDGA has values from 0 to 63 dB. The low-light imaging gain is primarily controlled through ground command of the VDGA (Elvidge et al., 1999). Before being launched, the OLS is calibrated in a simulated space-like environment. During this time, the OLS is exposed to light sources whose irradiances are known. The digital numbers are recorded over the full range of VDGA. The telescope input illuminance is made equal to digital numbers at specific VDGA gain settings by the calibration data (Elvidge et al., 1999). The relationship between OLS VDGA gain setting and the observed range of radiances based on preflight sensor calibration is shown in figure 2.1.

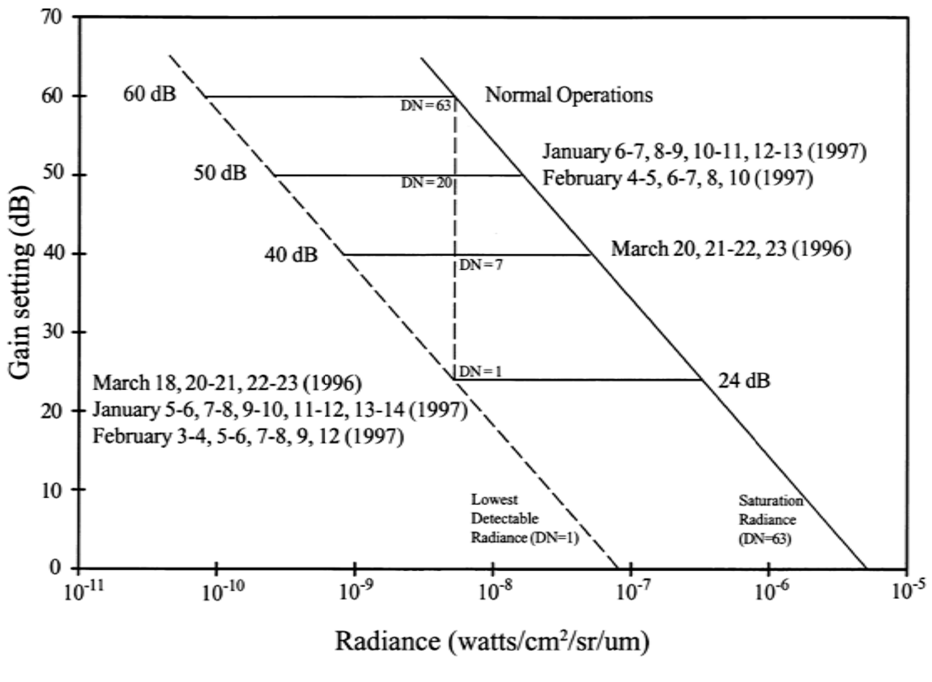


Figure 2.1: Relationship between F-12 OLS visible band gain when operated in PMT mode, digital numbers and observed radiances derived from the preflight sensor calibration (Elvidge et al., 1999 p. 79)

The solid diagonal line in the figure is the saturation radiance. Pixel values with observed radiances greater than saturation radiance have digital number values of 63. The dashed line is the lowest detectable radiance and corresponds to a digital number value of 1. At VDGA gain setting of 63, the total system gain is 136 db which is about $10^{13.6}$ times intensification of light (Elvidge et al., 1999).

This enables detection of radiance down to 10^{-10} watts/cm^2/sr/μm, (i.e. approximately 6 orders of magnitude) lower than the OLS daytime bands, Landsat TM or NOAA AVHRR.

In the OLS sensor, the gains of the visible bands are computed on board based on lunar phase and solar and lunar elevation (Elvidge et al., 1997b). An on board along scan algorithm modifies the base gain every 0.4 milliseconds. Also there is a BRDF (Bi-directional Reflectance Distribution Function) that enhances specular reflectance by adjusting the gain in the scan segment where the illumination angle approaches the observation angle (Elvidge et al., 1997b, Elvidge et al., 1999). Due to these two types of along scan gain changes visually consistent imagery of clouds are produced at all scan angles for use by meteorologists in the Department of Defense. In normal operations, scene illumination from lunar phase and elevation are tracked by the VDGA (Elvidge et al., 1999). During the darkest twelve nights of the lunar cycle, the visible band cannot record due to the low illumination. During this time, the along scan gain and the BRDF are minimized and the gain reaches its maximum. Gain settings gradually rise as lunar illumination declines reaching near 60 dB. The lowest gain settings occur under full moon conditions when the images produced look very similar to the daytime images recorded by the DMSP-OLS sensor.

2.1.3: DMSP-OLS Images:

The initial purpose of launching the DMSP-OLS programme was to record cloud cover and cloud top temperatures. However, in late 1970s it was discovered that the data obtained from this sensor was capable of capturing the artificial lights from cities at night (Croft, 1978, Welch and Zupko, 1980). The DMSP-OLS data was stored initially in analogue format or film strips till 1992 in National Snow and Ice Data Centre (NSIDC). This film strip data was used by Sullivan (1989) to produce a 10 Km resolution global image of visible and VNIR (Visible Near Infrared) emissions from the earth surface as observed by the OLS sensor. Sullivan (1989) produced global maps from OLS data captured on single dates. Cloud free images were selected and the VNIR emission sources were detected and mosaicked into a global product. As a result, the emissions from fire in places such as Africa were mapped in the map. The digital archive of DMSP-OLS was available from 1992. The availability of the digital data enhanced the scope of application of DMSP-OLS in more researches.

There are three types of datasets obtained from the DMSP-OLS satellite (Doll, 2008).

1. The frequency detection or the stable lights data;
2. Radiance calibrated data;
3. Average DN data.

2.1.3.1: The Stable lights dataset:

The city-lights dataset was produced from smooth resolution OLS data (2.7 Km) from 236 orbits (50 from F10 satellite and 186 from F12 satellite) for the period of October 1994 through March 1995. This dataset was prepared by Christopher Elvidge and his team in NGDC/NOAA in Boulder, Colorado, United States. This may be regarded as the "single greatest advance in the processing of OLS night-time light data" (Doll, 2008 p 8). The data was chosen from nights with low lunar illumination (less than half moon) covering 20 to 55 degrees north latitude. During these nights, the VNIR band gain on the OLS were set to their highest, as a result smaller light sources present on the earth surface could be detected. Lunar illumination posed a hindrance in image processing by reducing the contrast of lights from the ground. (Elvidge et al., 2001b). Another hindrance was the glare of the scattered sunlight (an adverse effect of the light intensification by the Photo Multiplier Tube) (Elvidge et al., 1997b). Glare (defined for DMSP-OLS dataset as 40 by 40 block of pixels with all pixels having saturated Digital Number values of 63) was removed. Lightning was removed due to the removal of the bad scan lines (defined as 10 consecutive lights with no lights above or below in DMSP-OLS data) (Small et al., 2005, Doll, 2008). The cloud screening was based on Thermal Infrared (TIR) band of the OLS sensor. The data was segmented into five latitudinal bands and a TIR threshold was visually selected for the images. Cloud pixels in each orbit were geolocated and projected into a reference grid. In the grid, the cell values increased by one for each orbit in which a cloud was present. This was followed by a calculation of the total number of cloud free observations in the reference grid. It was determined by subtracting the cloud counts from the coverage counts present in the grid Brightness variations within and between orbits prevented the selection of a single digital number threshold to identify the VNIR emission sources. An algorithm was established for automatic detection of the VNIR light sources using a nested configuration of 20 by 20 and 50 by 50 pixel blocks. This algorithm applied a threshold to the central 20 by 20 pixel block with the help of the background information obtained from the surrounding 50 by 50 block. There was a 60% overlap between the 50 by 50 pixel block which were used to generate the background statistics. This also reduced the threshold disparities between adjacent 20 by 20 pixel blocks resulting in smooth transitions. The lower limit of the background was selected as Digital Number (DN) = 1 (as DN = 0 is no data pixel). The upper limit of the threshold was calculated using a frequency distribution of pixel counts against DN values. Background pixels were determined by calculating the frequency distribution starting from the brightest pixel (DN = 63) to identify the first pixel value where five consecutive DN values had greater than 0.4% (10 pixels for a full 50 by 50 cell) of the total pixel counts. The threshold was established by calculating the mean plus four standard deviations of the background pixels (Doll, 2008, Elvidge et al., 1997b). A reference grid plane was then generated with the georeferenced pixels having cloud free VNIR emission sources. The number of times a cloud free light was detected in each cell was divided by the total number of cloud free observations and multiplied by 100 to generate the percent frequency dataset of cloud free OLS observations in each grid cell.

The stable light dataset was used to identify a number of light sources. It could be filtered to produce a number of different products such as products identifying lights from human settlements and industrial facility (city lights); fires; gas flaring and shipping fleets (Elvidge et al., 2001b). The city lights were identified by their stability over time period while the lights from forest fires could be identified due to their location and temporal instability. For preparing data showing the different light sources, there was a change in the original algorithm used for preparing the frequency distribution or the stable lights dataset. The image cell used to identify the background threshold was changed from 50 by 50 to 100 by 100 pixels. The original 50 by 50 pixel block corresponded to a ground distance of approximately 135 Km by 135 Km which proved inadequate to identify larger areas of lighting such as the Nile Delta. The other major difference was the use of 30 arc second grid instead of the Goode Homolosine Projection 1 Km equal area grid. The percent frequency dataset was filtered into a number of data products by manual editing and threshold filtering to separate lights into four categories: fires, fishing boats, gas flares and human settlements. Fires were extracted on the basis of their low frequency of occurrence, diffuse spatial pattern and location on the land surface. Only pixels with values greater than one were included to avoid the inclusion of noise. In the six months period analyzed (October 1994 to March 1995), fires were identified in Sahara desert, Southeast Asia and northern South America. The presence of cloud cover in the tropical regions posed a problem for detection of fires from those latitudes. Also, there were few observations of fire in Brazil and southern hemisphere as the analysis window of the satellite was optimized for the northern hemisphere and missed the primary burning season in these areas. Fishing boats were located in the similar manner to the fires, except that their location was on water. Large fleets of fishing boats were identified off the shores of Japan, parts of Asia such as Gulf of Thailand and off the Philippines and the coast of Argentina. Gas flares are identified by their lack of association with populated places and large circular appearance. Offshore gas flares were associated with heavily lit oil and gas platforms. All the remaining stable lights were identified as human settlements (Elvidge et al., 2001b).

A more recent stable light dataset was produced by NGDC spanning the years 1992 – 2009 (Baugh et al., 2010). There was an improvement in the dataset over the previous stable light products for identifying areas of persistent lighting. These products were created by compositing the cloud-free orbital sections in the centre half of the swath, with no sun and moonlight present. This was followed by a process of filtering of noise and transient light sources such as fires. The grid cell size of these images was 30 arc second or 1 square kilometre at the equator. However, the bright areas such as urban cores were saturated in these images and variations in distribution of radiance within these areas were indiscernible. In order to overcome this problem, radiance calibrated products were prepared. The details of these products are described in the following section.

2.1.3.2: The Radiance calibrated product:

This dataset was produced in 1999 by Elvidge et al (1999) to overcome the problem of relatively low light pixel saturation in stable lights dataset and to detect a greater range of settlements by varying the gain of the sensor. Experiments were carried out in which data was collected on the darkest nights on March 1996, starting from 16^{th} March with the ASGC (Along Scan Gain Control) and BRDF functions of the OLS sensor turned off. Two gain settings of 24 db and 50 db were selected. They were alternately applied to collect data between 5-14 January and 3-12 February in 1997 (the darkest nights in January and February of 1997). The data was then processed through the primary steps such as establishment of a reference grid, identification and geolocation of light sources, clouds and coverage areas; a scale to convert digital number to radiance, compositing cloud free portions within two overlapping gain ranges (high and low); thresholding to eliminate the isolated detections and finally combining the calibrated images by averaging based on the number of detections (Elvidge et al., 1999). The algorithm used to prepare the stable lights product was not suitable to determine the threshold in the radiance calibrated product as that performed poorly in identifying diffuse lighting which was often dim and spatially scattered across the landscape. A software tool was therefore developed and used to identify areas of diffuse lighting as well as the focused lights from cities and towns. VNIR emission sources were identified using the software. Clouds were detected using the OLS thermal bands. Each suborbit was divided into 200 pixel-wide latitudinal bands. Thermal infrared thresholds for the identification of cloud pixels were visually carried out. Geolocation of lights, clouds and coverage polygons were done using an algorithm which estimated the latitude and longitude of pixel centres based on geodetic subtrack of the satellite orbit (modeled by the Air Force orbital mechanics models that uses the daily radar bevel vector sightings of the satellite as input to calculate the satellite position every 0.4208 sec); satellite altitude; OLS scan angle equations, an Earth sea level model and digital terrain data (Elvidge et al., 1999). The final product was produced by averaging the high and low gain cloud free composites. The conversion of DN to radiance was given by the formula:

$$\mathit{Radiance} = (\mathit{Digital\ Number})^{3/2}\ X\ 10^{-10}\ \mathit{Watts/cm^2/sr/\mu m}$$

Equation 2.1: Conversion of DN to radiance

The sub-orbits were divided into two groups of data, one acquired in a gain setting higher than 30 dB and the other with a gain setting of lower than 30 dB. The cloud free observations were calculated for each of these data groups. Average radiance for each reference grid pixel was calculated in the final stage of producing the composite image. All noises and ephemeral lights detected once in both the high and low gain images were removed in the next stage and finally the high and low gain cloud free composites were averaged (Elvidge et al., 1999).

The radiance in the data product ranged from $1.54X10^{-9}$ Watts/cm^2/sr/μm to $3.17X10^{-7}$ Watts/cm^2/sr/μm. The range of radiance was made considerably ample on both sides to accommodate any future change in the gain setting. This data product was different in character than the other DMSP-OLS images. This data provide details of brightness variations within urban centres. It also showed diffuse lighting in the densely populated suburban and rural environments. As this dataset showed variation in the DN values, which is physically more meaningful than lit-frequency (Doll, 2008), this data proved more useful in modeling and finding appropriate relationships between parameters of interest.

A recent global composite radiance calibrated data was prepared for 2006 by combining the low and high gain setting datasets (Ziskin et al., 2010). This dataset was prepared with no sensor saturation using the low and high gain datasets. In order to overcome the dynamic range limitations of the OLS, a stable light dataset of 2006 was merged with a fixed gain dataset produced using high, medium and low gain data. The fixed gain products (low, medium and high) were integrated by a process of simple merge modified by a weighing factor (Ziskin et al., 2010). The weighing factor was calculated on the basis of the fall in DN value relative to its overlap with the range of other fixed gains. The problem of saturation of bright lights was overcome after the fixed gain product was merged with the stable light dataset to produce the radiance calibrated data.

2.1.3.3: Average Digital Number:

This is the latest and the most extensive night-time lights data. This was originally launched as the change product and offered two global coverages (for 1992-93 and 2000) and a change product.

This data was processed from high quality visible band DMSP-OLS data with no lunar illuminance, glare, marginal data, bad scan lines or lightning (Sutton, 2003a). Pixels with solar elevation angle of more than six degrees were flagged as sunlit. This data was used to eliminate glare which was described as a condition caused when sunlight is scattered into the OLS telescope. Glare was detected where 75 or more saturated band pixels were found in a 40 by 40 tiled window. Marginal data were located poleward between the glare and sunlit data and were detected by slightly elevated DN when compared with high quality night-time light datasets. The bad scan lines with ten consecutive pixels with no light above or below them were also removed. 30 arc second grids were created by a geolocation algorithm which operates in the forward mode projecting the centre of each pixel on the surface of the earth. Cloud detection was achieved using the thermal band data. An automatic cloud detection algorithm was used. It was based on differencing the OLS thermal band brightness temperatures with gridded surface temperature values produced by NOAA National Centre for Environmental Prediction (NCEP). The NCEP surface temperature grid was resampled to 30 arc second grid and each of the OLS processed image was temporally matched with this grid. The problem of accurately detecting clouds along coastal boundaries was solved using a look-up table which assigns surface temperature to 30arc second OLS grid cells from the 1 degree and 2.5 degree NCEP cells lying on both land and oceans. A temperature difference image was created. A temperature difference histogram showed bell shaped peak at or near the origin with the maximum peak around 15 degrees from the origin. The cloud detection threshold was calculated as mean plus two standard deviations of the Gaussian with the difference threshold lying between five and twenty degrees (Sutton, 2003a). The centre halves of orbital coverages were used in these images instead of the onboard BRDF correction algorithm. This consistent use of the imagery from the centre halves prevented large differences in the brightness values due to variations in scan angle. The blooming effects were manually adjusted by NGDC/NOAA (Doll, 2008).

This data was recently used to create a full archive from every sensor for every year. Elvidge et al (2007a) identified the main types of changes detected in the time series:

A. *Expansion of lighting around urban areas:* The expansion of development at the outer rims of many cities led to the increase in lighted areas in those parts. The saturation in the city centres prevented detection of changes in those areas. The detection of light change around the urban areas was detected by overlaying images of different years in red, green, blue (RGB) colour composite, by differencing the urban composites or by creating a transect from the city centre.
B. *Areas of new lighting:* New areas of light were detected in a number of areas for example in parts of South Africa and United Arab Emirates due to the recent developments in infrastructure and construction in those areas.
C. *Temporary lighting:* There were many occurrences of temporary lightings which appear and then disappear from the time series for example areas associated with mines. It indicated decommission of some facilities which might have been constructed for some particular purpose. Some linear temporary features were noted in China along lines of transportation development where people work at night.
D. *Loss of lighting:* There were areas like some places in Russia (for example Moldova) where there is a loss of light. In some other parts, there was return of the lights lost indicating a reconstruction of facilities in those areas.
E. *Reductions in skyglow:* Skyglows could be observed offshore of many big cities due to the ability of OLS to detect lights to 10^{-9} watts/cm^2/sr. The reductions in skyglows were

attributed to changes in lighting types or the installation of fixtures designed to direct lights downward. Offshore areas around Los Angeles showed a decline in the skyglow.

The average DN data product was available for in near global coverages expanding from -180 to 180 degrees and -65 to 65 degrees north. The four F-satellites on board the DMSP satellites collect data and there were six sets of co-temporal data to be compared between the satellites. The satellites and their co-temporal years are shown in the following table (table: 2.3):

Table 2.3: Co-temporal years between DMSP-OLS satellites

Satellites	Years	Co-temporal years with next series
F10	1992-1994	1994
F12	1994-1999	1997,1998,1999
F14	1997-2002	2001, 2002
F15	2001-2003	-

2.2: Census

The word census had its origin in the Latin word "censere" which meant to assess, to rate and to estimate (Maheshwari, 1996). Census is commonly defined as the counting of population of a country. It is more than a simple head-count of people and involves a complex collection of a wide range of information about each individual such as ethnic composition (tribes), education (literacy rate) and employment. A census categorizes the responses of people to make the data more manageable (Christopher, 2008). The Office of the Registrar General and Census Commissioner of India described population census as "*…the total process of collecting, compiling, analyzing or otherwise disseminating demographic, economic and social data pertaining, at a specific time, to all persons of a country or a well-defined part of a country. As such, the census provided snapshot of the country's population and housing at a given point of time*" (Office of the Registrar General and Census Commisioner of India, 2001a). The census is often regarded as a "view of society" whereby it is an attempt by the state to assess the number and characteristics of the population (Christopher, 2008, Scott, 1999).

The history of the census dates back to the rulers of ancient Babylonia, China, Egypt and Rome when information on people was collected to determine wealthy population who could be taxed and also who could work in the army (Encyclopaedia Britannica, 2007, Office for National Statistics, 2008). The census data was used for maintaining national identity (Anderson, 1999) and to define the national territory (Christopher, 2008). The information helped in taxation, prevention of crime, recording information on the work force and also protecting property rights and provision of welfare (Higgs, 2004). Although in the initial phase, absence of information technology made the job of making census labourious, information was collected because it was important and valuable. With the development of electronic means and introduction of database, census data could be easily integrated and updated (Higgs, 2004). The use of statistics enabled the state to gain greater role in deriving the necessary inferences from the census databases (Beaud and Prévost, 2005, Desrosières et al., 2006, Christopher, 2008).

2.2.1: History of the Indian Census:

There was absence of any reliable data on India’s population in ancient times. Kautilya’s Arthasastra written in 321 – 296 BC talked about the census in its Chapter 35 (Maheshwari, 1996, Shamasastry, 1915). The census was important in those days in state policy for the purpose of taxation. There were other historical documents that gave some account of the population in India. The Indus Valley Civilization (from 3000 to 1500 BC) had a significant urban population (Office of the Registrar General and Census Commisioner of India, 2001b). It was followed by the Aryan civilization (1500 – 600 BC) and recorded growing population (Office of the Registrar General and Census Commisioner of India, 2001b, Maheshwari, 1996). Ancient Indian literature such as the Rigveda accounted for scanty population living in villages (Maheshwari, 1996, Office of the Registrar General and Census Commisioner of India, 2001b). The epics of Ramayana mentioned uninhabited countrysides while the Mahabharata described populous lands (Maheshwari, 1996).

Before its independence in 1947, India was a colonial state under the British government. The government, of 1865, ordered a formal census to be conducted for India at that time. The purpose was to obtain estimates on the population of the country. However, the project was delayed because of the mutiny of 1857. The first Indian census was carried out in 1865 in the north western provinces (presently Uttar Pradesh and Uttarakhand), the Central Provinces and Berar in 1866, Bihar in 1867 and Punjab in January 1868. The cities of Madras (Chennai), Bombay (Mumbai) and Calcutta (Kolkata) had their first censuses in 1863, 1864 and 1866 respectively. Between 1867 and 1872, a series of censuses were carried out in the country based on a uniform set of schedules, but these were not centrally supervised and did not cover the whole of India.

The first regular census of India was taken in 1881. Since then census in India has been carried out every 10 years. This included a census in 1941 during World War II and also in 1921 during the Non-Co-operation Movement under Mahatma Gandhi (Maheshwari, 1996).

2.2.2: Hierarchy of regions for census enumeration:

India is divided administratively into states and union territories (the areas governed by the Central government directly and do not have separate state governments). The states are divided into districts, with each district divided into smaller units by the district collector (an officer, appointed by the Central Government of India, who is in charge of a district for the purpose of census enumeration). Each district is further subdivided into smaller spatial units consisting of, in decreasing order of area, charges, circles and blocks. This is shown in figure 2.2.

A block consists of 30 – 50 houses and is under the charge of an enumerator. The census schedule for the block is filled in by the enumerator from the information collected. The information is not published until the census night. On the census night the final enumeration by addition or omission as necessary is done by the enumerator. There is also a method to count the mobile population. The railway stations keep records of the people leaving on census nights and temporary residential services such as hotels, motels, tourist lodges, youth hostels and guest houses keep records of the place of origin and destinations of the people checking in and checking out on the census night. The houseless population such as cave dwellers and vagrants are also recorded by the Indian census, a process conducted since 1881 (Maheshwari, 1996).

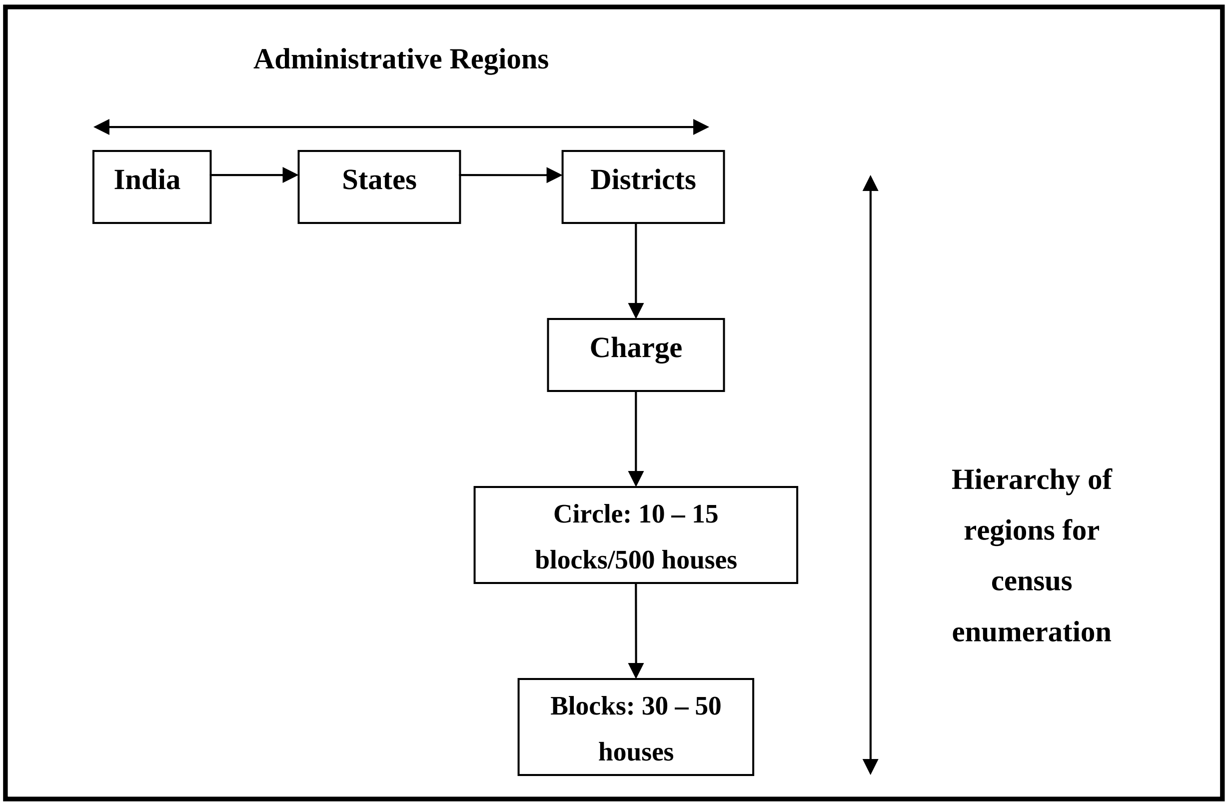


Figure 2.2: Hierarchy of regions for census enumeration

2.2.3 Census Personnel:

The Census Office in India is headed by the Census Commissioner of India. Other officials include Superintendent of Census Operations, District Officers, Subdistrict or Tehsil officers, and the Village Patwari (a land-record officer at the sub-district or village level).

Census in India depends heavily on the enumerators. The necessary qualification for the enumerators is the ability to read and write. Attempts to employ police force and volunteers as enumerators have been unsuccessful (Maheshwari, 1996). As a result, supervisors are usually selected from government departments such as the land revenue, public works, civil courts, forest surveys, settlements, educations, medicine and courts of wards (Maheshwari, 1996). These officials are considered amenable to government disciplines. Also recruitment of village patwaris and district collectors in the census process reduced the cost of carrying out enumeration efficiently.

2.2.4: Method of enumeration:

From 1881 to 1931 there was the system of headcount on the census night. Initially the census night was selected as a full moon night with no big festival. One-night enumeration necessitated the employment of a large number of enumerators. During 1931 census around two million enumerators were recruited. From 1941 census, the one-night theory of population enumeration was abandoned. It was modified to spread over a period, half on either side of the central census date. The census method changed from de facto to de jure in 1941 (Maheshwari, 1996). The change had many associated advantages such as increase in block size and decrease in the number of enumerators. From then on, the census recorded the number of people ‘normally’ located in an area along with all the nomads of the area who did not have any fixed dwelling (Maheshwari, 1996).

Initially the enumerators collected information on age, sex, caste and occupation. The information obtained on age, caste and occupation was incomplete and unreliable in many occasions. This was mainly because of the lack of information on these areas with the people. For example, sometimes the enumerators themselves were not aware of the ages and castes of the target population. Indian census has been conducted uninterruptedly from 1881. It is one of the most extensive population counting activities in the world to date. The last complete Indian census was in 2001. Details of the 2001 census are given in a later section (section 2.3).

Population enumeration for the 2011 census in India was conducted in February and March of the same year. Provisional population totals have been published. This data was not used in this research because of two reasons. Firstly when the research commenced in 2008, the data from 2001 census was the latest available dataset. Secondly, the complete census of 2011 is not available to date. The provisional tables are presented by adding up the population data collected by each census enumerator and suffers from errors in calculation and duplication or omission of enumeration units (Office of the Registrar General and Census Commissioner, 2011a).

2.2.5: Collection, Assimilation and Publication of the census:

The collection and assimilation of census data are undertaken by the Data Processing Division in the Office of the Registrar General, India (ORGI). The technology used to record the census data has improved with time. Until 1961 census, data collection, data entry and processing were done manually. The data schedules were prepared at various regional tabulation centres throughout the country and these were then sent for data entry. Data processing was done on a sample selected from the entire data collected. Data were converted to machine readable form by using hand punching machines (inserting one card at a time) using 80 column (Hollerith) punch cards (Office of the Registrar General and Census Commisioner of India, 2001a). Serial Rolling Total Tabulator cum printer (SRTT) was used for tabulation and printing of census tables. There was no system of data backup at that time.

The technology used in the 1971 census was a little improvement on that of the previous one. The hand punching machine used in this census was capable of using 80 column (Hollerith) punch cards together instead of inserting one card at a time. Also, an IBM 1401 computer along with an IBM card reader and printer were used at the ORGI Office in New Delhi for data processing and printing. Spools of magnetic tapes were used for data storage.

During 1981 census, data entry activities were carried out in 15 data centres set up in major states along with one at the headquarter at ORGI. Each centre had a number of adjoining states under its jurisdiction for data entry using “key to disk machines” (Office of the Registrar General and Census Commisioner of India, 2001a). The data processing was carried out at the National Informatics Centre (NIC), New Delhi and Regional Computer Centre (RCC), Chandigarh.

During the 1991 census, there was an important change in the data processing activities. A Medha – 930 main frame computing facility was installed at the Data Processing (DP) Division at ORGI. 163 data tabulation centres, 15 data centres, 4 Regional Processing Centres (Delhi, Bhopal, Bhubaneswar and Chennai) were set up during the 1991 census to divide the whole task of data entry and tabulation. The data schedules were coded at the 163 data tabulation centres distributed all over the country and the coded schedules were sent to the 15 data centres for data entry at their “Dump Terminals” (Office of the Registrar General and Census Commisioner of India, 2001a). A UNIX operating system connected the main server at ORGI with those at the other 15 data centres. Master data files creation, data editing on basic fields and lower level tabulations were processed at the 4 Regional Processing Centres although the major editing on all the fields, processing and generation of tables at various levels were taken up by the DP Division at ORGI. Processing was done on 100% of the data of workers and Scheduled Caste/Scheduled Tribe (SC/ST) while 10% of the other records were processed.

2.3: Indian Census 2001:

The last Indian census was carried out in 2001. The 2001 Indian census was the first population counting activity of the 21st century. Unlike previous censuses where a sample of data collected was processed to prepare the final census, the 2001 census provided results from 100% of the data collected. More than one billion records were processed before the final results were published. It was predicted by the ORGI that the results from the 2001 census would give an insight into the effects of the liberalization and globalization of the Indian economy, which was opened to the global market in 1991. This was the 14th census since 1872 and the sixth census after independence (in 1947). More than 2 million enumerators were deployed during the process of census taking in 2001. Every household in the country was visited to collect information. There was large scale hardware up-grade at the 15 data centres and the DP Division at ORGI. Intelligent Character Recognition (ICR) technology was introduced for the first time. Around 228 million paper forms were scanned for the first time in census history and the scanned images were saved for further reference (Office of the Registrar General and Census Commisioner of India, 2001a). In all these respects the 2001 census was a major improvement over the previous censuses carried out in India.

The census was carried out in two phases. The first phase was the House – Listing Operations and the second phase was the Population Enumeration. The details of data recorded in the House – Listing operations and the Population Enumeration are given in table 2.4.

Table 2.4: Parameters considered in House-Listing operations and Population enumeration surveys in the *census*

House – Listing Operations	Population Enumeration
• Use of the census houses • Condition of census houses used as residence • Predominant material of the roof, wall and floor or the census houses • Type of structure of census houses	***For each member of the household :*** • Relation to head of the household • Sex • Age at last birth day • Current marital status

• Number of dwelling rooms • Ownership status of the house • Number of married couples and whether they have independent sleeping rooms • Source of drinking water (e.g., Tap; Hand pump; Tube well; Well; Tank; Pond River; canal; Spring; Other) and its location • Source of lighting (e.g., Electricity; kerosene; Solar energy; Other oil, Any other; No lighting) • Availability of bathroom, type of latrine and type of drainage for waste water • Availability of separate kitchen and type of fuel used for cooing (e.g., Firewood; Crop residue; Cow dung cake; Coal, Lignite, Charcoal; kerosene; LPG; Electricity; Biogas; Other) • Availing of banking services and availability of the specified assets (e.g., Radio, Transistor; Television; Bicycle; Motor Cycle, Moped; Car, Jeep, Van; None of these	• Age at marriage • Religion • Name of Scheduled Caste / Scheduled Tribe • Mother tongue • Other languages Known • Highest educational level attained • If attending educational institution • Disability status • Worker / non –worker • Main Worker / Marginal Worker • Economic activity of the main or marginal worker • Non –economic activity of Marginal worker and non –worker • Marginal worker or non-worker – seeking / available for work • Distance and mode of travel to place of work • Place of Birth • Place of last residence • Duration of residence at place of enumeration • Reason for migration ***For All Married Women:*** • Number of children ever born • Number of children surviving ***For Currently Married Women:*** • Births during previous year ***For cultivation households*** • Area of land cultivated

	• Tenancy Status

2.3.1: Summary of results as obtained from the 2001 census:

At the time of census 2001, administratively India was divided into 28 states and 7 Union Territories. The total area of the country was 3.2 million Km^2. The largest state in India was Rajasthan with a total area of 342,000 Km^2. Each state and union territory was divided into districts. There were 593 districts in India in 2001. Each district was divided administratively into sub – districts which were known by various names such as Tehsil, taluk, Community Development (CD) block, Police Station, Mandal and Revenue Circle in different parts of the country. The lowest administrative division of the country was a village in rural area and town in urban area. These are the administrative divisions in the country and are different from the census enumeration zones as described in section 2.2.2. During 2001 there were around 639,000 villages in the country of which approximately 45% (44,856 villages) were uninhabited. The population size of a village depended on its location and availability of land and water. For example, villages located on the plains and beside big rivers were more populated than those located in the mountainous areas. The Indian census also accounted for Urban Agglomerations (UA). These were the areas adjacent to notified towns which acquired an urban character. According to the Indian census there were 384 UAs in 2001. Sometimes, outgrowths of bigger towns were also regarded as urban agglomerations. These were also regarded as UAs/towns. There were 4378 such UAs/towns in 2001. Although the towns and villages formed part of a subdistrict, there were towns which extended beyond the boundary of a district such as Delhi and Srinagar. Some of the districts in the census 2001 were completely urban in character such as districts of Mumbai, Chennai, and Kolkata were completely urban with no rural population. Also there were some districts which housed only rural population.

At 00:00 hrs on 1st March 2001, the total population of India was approximately 1.03 billion. The state of Uttar Pradesh was the most populous state with a total population of 166 million. It was followed by the states of Maharashtra (97 million) and Bihar (83 million). The Union Territory of Delhi was the most populous with 13.8 million inhabitants. Most of the villages in India at that time had a population of 500 to 999 people. However, there were 3,961 villages with a population of 10,000 or more. There were 1346 UAs/towns in India in 2001 with a population ranging from 10,000 to 19,999. The largest UA was Mumbai with a population of 16.04 million followed by Kolkata (13.2 million) and Delhi (12.9 million). There were 35 UAs/towns with more than one million people.

Other demographic profiles as obtained from the census 2001 are presented in table 2.5.

Table 2.5: Selected demographic profiles as obtained from Indian census 2001

Demographic variables	Results
Gender ratio (number of females per 1000 males)	933
Child gender ratio (0 – 6 years)	927
Workers	39.1% of total population (51.7% of total males and 25.6% of total females)
Population aged less than 18 years	41% of total population
Literacy rate	64.8%
Highest number of literates (both rural and urban)	Kerala
Scheduled caste	16.2% of total population
Scheduled Tribe	8.2% of total population
Gender ratio of 1000 and above (i.e. more than 1000 females per 1000 males)	558 UAs/towns 174,351 villages
75% or more literate population	2516 UAs/towns 93,055 villages

The census also collects information on households and the different household amenities. A summary on the households and household amenities as obtained by the census 2001 are given in table 2.6.

Table 2.6: Selected household amenities data as obtained from the Indian census

Household Amenities	Results
Total number of households used solely for residential purpose	187 million
Houses made of concrete	19.8% of total households
Households using electricity as source of power	55.8% of total households
Television available	31.6% of households
Available bank facilities	35.5% of households

2.4: Application of Remote Sensing in census and population studies:

The growth in population in any settlement is mostly accounted for by censuses in most of the countries. These censuses however lack information on the areal extent of urban areas. Also, apart from a few large cities, there is no reliable data on the location of the urbanites (Potere and Schneider, 2007). In order to look into these issues, attempts were made to estimate population from satellite images which would enable mapping the location and areal extents of the settlements. The application of remote sensing in estimating socio-economic data including census is described in the following sections.

2.4.1: Association of remote sensing and socio-economic data:

Remote sensing techniques have been used to collect census and population data since the 1990s (Liu, 2003, Lu et al., 2010, Lo, 2001a, Lo, 1995, Wu et al., 2008, Lo, 2008b, Lu et al., 2006). The method of predicting social and human phenomena from remotely sensed images was referred to as indirect interpretation (Massey, 2002). Remote sensing datasets are capable of providing synoptic views and repeat coverages of study areas. This has facilitated the application of this technique for predicting population or census (Lo, 2006). The first proposal for using remotely sensed images in census estimation for the United States was put forward by Brugioni (1983). High resolution aerial photographs were used at that time to estimate population on the basis of the number of houses counted from the photos. Similarly Morgan (1984) used aerial photographs along with demographic sample ground surveys to estimate population of Lagos Metropolitan Area in Nigeria.

There were different approaches to population estimation using remotely sensed images. These included identifying the number of dwelling units (Lo, 1995), classifying landuse types such as built up areas and correlating spectral reflectance to population (Lu et al., 2010), studying population dynamics (Morton and Yuan, 2009), extracting impervious surface (Lo, 2008a) and Dasymetric Mapping Principle (Yuan et al., 1997, Martin, 1992, de Sherbinin et al., 2002). For example, in the United Kingdom, census data was linked with urban land use classification from SPOT and Landsat images of Bristol, Swindon, Norwich and Peterborough (Mesev, 1998). The population zone data was overlaid on residential areas obtained from classified Landsat image. Socio-economic metrics were interpolated into raster based surfaces using a Geographic Information System (GIS) (Martin and Bracken, 1993). Martin (1989) incorporated the population from centroids to a raster grid on the basis of local data point density for each enumeration district of United Kingdom. However, the reliability of the approach was dependent on the accuracy of the residential class obtained from the classified satellite imagery.

Data from Landsat (TM and ETM+) were widely used also for studying socio-economic aspects. These images were used in estimating population density models as well as in association with census data. Residential population density models were estimated in Marion County, Indiana using impervious surface approach from Landsat ETM+ data (Lu et al., 2006). Urban growth of twenty-five mid sized cities from around the globe with population ranging from 1 million to 5 million was compared using remotely sensed images and census data (Schneider and Woodcock, 2008). The cities had different geographical settings and levels of economic development. The work aimed at better understanding of the urban growth of cities around the globe in the decade 1990-2000.

Changes in urban built up area and population distribution pattern using Landsat TM and ETM+ images, fieldwork and images from CORONA, IRS and IKONOS satellites were studied for Greater Cairo, Egypt (Yin et al., 2005). Analysis against census data from 1986 and 1996 showed that population per unit area of built up surface was a good indicator of urbanization. Census data along with Landsat ETM+ images were also used to estimate population density of Indianapolis, Indiana (Li and Weng, 2005). Spectral signatures, principal components, vegetation indices, image textures and temperature were correlated with population.

2.4.2: Estimation of intra-settlement population:

Remote sensing techniques were extended to estimate population within settlements. Images captured from both aerial and satellite platforms were used to estimate population within settlements. Aerial photos were used to estimate population within settlements (Lo, 1979). The number of households was counted from the aerial photos and multiplied with population within each household. The household size data was obtained either from the census or from field surveys. Mathematical models relating population to landuse studies were used to estimate population from aerial photos. This approach was first introduced by Kraus (Lo, 2006) who used panchromatic photographs and colour infrared photographs to estimate population of four cities of California. Later it was used in other places such as Nigeria (Hsu, 1973) and Hong Kong (Iisaka and Hegedus, 1982).

Population of settlements were also estimated from satellite images such as Landsat MSS and SPOT (Lo, 1995). Landsat MSS images were first used to predict population by Hsu (1991) and was later expanded for predicting population of Tokyo (Harvey, 2002). It was assumed in the study that areas with low *population density* were covered by more natural landcover such as vegetation and bare soil Chen (2002) used 20m SPOT multispectral data to estimate the population of Kowloon peninsula of Hong Kong. Tertiary Population Units were used to aggregate population obtained from 1986 by-census population data. Langford (1996) used Landsat TM images of northern Leicestershire to estimate population in a 1Km by 1 Km grid with the 1981 population census at the ward level. Landsat images were also used to estimate population of census districts in Australia. Harvey (2002) estimated population counts and densities for Ballarat and Geelong Statistical Districts. Significant correlations were obtained and urban population was estimated with an error of 1% for training set data. The results were more accurate for the suburban areas with moderate population densities. High density areas were underestimated while low density areas were overestimated. Micro, medium and macro levels of landuse classification from Landsat TM images was linked with census data on hierarchical census districts for Australia (Webster, 1996). In addition to images from Landsat, data from Quick Bird satellite and LIDAR were fused to predict residential population at census blocks in Denver, Colorado (Lu et al., 2010). Both area based and volume based approaches were used. Residential building footprints were delineated. Regression models were used to correlate residential population with area and volume of buildings.

Homogeneity of residential areas in census districts were sometimes measured using texture statistic. It was observed that census data had higher correlations with areas of different residential densities than with an aggregated whole residential area. Urban texture analyzes to detect morphological patterns and improve the accuracy of population and dwelling estimation was proposed by Webster (as cited in Lo, 2006). Texture was defined as the "frequency of change in tones" (Geoghegan et al., 1998) and is a variation of colour or shades which is scale-dependent. Webster (as cited in Turner, 1990) used the dispersion of a single pixel value from an average (variance or standard deviation) as a measure of texture. SPOT panchromatic data was used to estimate number of dwellings in Welsh, Cardiff. It was concluded that advanced pattern recognition would improve the estimation of dwellings. It was also proposed that spectral radiance was better applicable for dwelling unit estimation than population estimation because satellite images more efficiently capture the reflectance characteristics from the dwellings (Geoghegan et al., 1998).

Satellite images were also extended to estimate socio-economic metrics. The application of remote sensing techniques in predicting socio-economic metrics are explained in the following section.

2.4.3: Estimating socio-economic variables:

Although socio-economic variables were not directly observed from remotely sensed images, indirect information of these variables could be obtained from aerial and satellite images. Geoghegan (1998) addressed the application of remote sensing to social science studies. It was noted that indicators of social and human environmental conditions were generated by many processes in place and were found in complex composite patterns in remotely sensed images. Geoghegan (1998) proposed that statistical techniques were essential for proper estimation of population from pixel. More emphasis on models such as the Markov Chain model (Metivier and McCoy, 1971, Henderson and Utano, 1975) and other spatial probability models were proposed. There was also need to address issues of scalar dynamics such as path dependence, buffering and amplification (Weber and Hirsch, 1992).

Use of remotely sensed images in social studies focused on the aspects of poverty, population density, income, education level and house characteristics. As early as the 1950s, large scale aerial photographs were used to study the social structure of American cities (Li and Weng, 2007). Later, aerial photographs were also used to assess housing quality and delineate the poverty stricken areas within cities (Lo, 1997, Lo and Faber, 1997). From the 1990s, satellite images were used to study socio-economic aspects of urban areas. In France, SPOT multispectral and panchromatic images were used in conjunction with census data to predict socio-economic variables for the city of Strasbourg (Walsh et al., 1999). In the US, greenness, impervious surface and temperature were derived from Landsat images of Indianapolis and correlated with socio economic metrics such as population density, income, poverty, employment rate, education level and house characteristics from the 2000 US census data (Lo, 2006).

Socio-economic variables such as total population, total land under cultivation, gender ratio and the number of households were correlated with environmental conditions such as slope angle, soil moisture potential obtained from Landsat TM data in Nang Rong District of Thailand (Lo, 2006). The results showed that ratio of total land under cultivation and total population was mostly associated with elevation and soil moisture content, total number of households were better correlated with the total biomass while gender ratio was mostly correlated with slope angle (Paul, 1984, Lo, 2006). Another important social application of remote sensing was the use of satellite images and census data to study quality of life (Lo, 2006). Quality of life was conceptualized as a collective attribute for a group of people living in similar morphological and socio cultural environment in a social space. Material welfare, environmental conditions and crowdedness were identified as important factors to assess quality of life. In Athens-Clarke County in Georgia landcover, vegetation index and surface temperature were obtained from Landsat TM images in order to assess the quality of urban life (Morrow-Jones and Watkins, 1984).

2.5: Applications of DMSP-OLS datasets:

DMSP-OLS data were used in studies of social sciences focusing on different areas. As these images are captured at night, they provide special advantages in detecting human habitation and activities from artificial sources of light. The following sections describe the application of the dataset in various applications. Important areas of application of DMSP-OLS data were its use in delineating and mapping urban areas, population studies, environmental applications and disaster management.

2.5.1: Application in mapping urban areas and settlements:

One of the most important phenomena affecting the world today is the process of urbanization. Broadly defined as the movement of people from rural to urban areas, it is prevalent in today's world with more than 50% of the global population living in urban areas (United Nations Human Settlements Programme., 2006). It has been predicted by the UNHabitat, that by 2030, 60% of the global population will be living in cities; of which 85% will be concentrated in the developing countries located today mainly in Asia and Latin America (United Nations Human Settlements Programme., 2006). It is important to map the existing and sprawling urban areas. It is also equally important to study the growth of new settlements. Artificial lights from areas of human habitation are recorded by DMSP-OLS images. The information obtained from these datasets has been used as surrogates in mapping urban areas for more than two decades. Georeferenced inventories of settlements of Europe, North and South America and Asia were created from the time series DMSP-OLS images (Elvidge et al., 1997a).

One of the most popular methods to map urban settlements from DMSP-OLS datasets was the use of thresholding techniques. This was applied for cities in United States (Imhoff et al., 1997b, Sutton et al., 2006), Brazilian Amazonia (Amaral et al., 2005) and India (Roychowdhury et al., 2009, Roychowdhury et al., 2011c). It was found that different thresholds were required for cities at different levels of economic development (Small et al., 2005). While a threshold of 89% was proposed for cities of United States (Imhoff et al., 1997b), a threshold of 30% was optimum for Brazilian cities (Amaral et al., 2005). Henderson (2003) applied thresholding for the three cities at different levels of urbanization and economic development such as Beijing, Lhasa and San Francisco. The boundaries derived were compared with the built up areas derived from Landsat TM data. Roychowdhury (2009) applied the method of thresholding to four cities of Maharashtra such as Mumbai, Pune, Nagpur and Nashik. The study showed that the method of thresholding was helpful to classify the cities on the basis of their population. Three distinct thresholds at 20%, 50% and 70% were identified as significant. Urban foot prints of Hyderabad derived from Landsat ETM+ data and DMSP-OLS image were compared (Roychowdhury et al., 2011c). The results showed an accuracy of more than 90% in the classification and delineation of urban, suburban and rural landcover types. It was demonstrated that in addition to spectral reflectance captured by satellites, differences in the degree of brightness of the lights emitted from urban areas at night was also an effective indicator in delineating landcover types.

DMSP-OLS and MODIS data were integrated to monitor growth of built up areas in the Indo – Gangetic regions of India (Roy Chowdhury and Maithani, 2010). DMSP-OLS data was used together with MODIS NDVI data to prepare a Human Settlement Index and built up area was estimated on a per – pixel basis. Another method to delineate urban areas using DMSP-OLS datasets was called “boosting”(Schneider et al., 2003). Urban areas at a global scale were mapped from DMSP-OLS dataset by this method ‘Boosting’ method was used to improve the accuracy of supervised classification and the result was fused with MODIS data and population data. Results from US suggested that the data fusion increased the accuracy of classification of land cover classes (Schneider et al., 2003). Cao et al (2009) proposed a Support Vector Machine (SVM) based algorithm using DMSP-OLS and SPOT NDVI datasets to extract urban areas semi – automatically.

In addition to mapping urban areas, the method of thresholding was also used to delineate urban sprawl (Sutton, 2003b, Qi and Chopping, 2007, Lo, 2001b, Lo, 2001a, Stewart and Warntz, 1958). Two radiance thresholds of 900 watts/cm^2/sr/μm and 2000 watts/cm^2/sr/μm were used respectively for identifying urban agglomerations and cities in the US (Sutton, 2003b). The DMSP-OLS datasets were applied to study urban areas in different areas in China such as those in the Yellow River region in inner Mongolia, over the period 1992-2003 (Lo, 2001b, Lo, 2001a, Qi and Chopping, 2007). The results showed an expansion of urban areas in China in recent past. In Bohai Rim area of China, the nature of urbanization was modeled into three types: polygon urbanization, line urbanization and point urbanization (Zhuo et al., 2009).

However, one of the drawbacks of estimating urban extents from DMSP-OLS dataset was the overestimation of urban areas. The DMSP-OLS data overestimated urban area by about 20 times when compared with the Digital Chart of the World dataset (Doll and Muller, 1999). Small (2005) applied thresholds to reduce the effect of over extent or “blooming” for seventeen cities. A 14% threshold was identified to reduce blooming around bigger cities for stable lights dataset of 2000. However, the threshold for bigger cities attenuated the smaller settlements.

Along with mapping urban areas by the method of threshold, DMSP-OLS data was also used to map impervious surfaces across the globe. Impervious surfaces contribute to urban heat island effects, altering the watershed characteristics by changing the frequency and magnitude of surface run-off, reduce carbon sequestering from the atmosphere (Milesi et al., 2003), increase flooding, reduce ground water recharge and adversely affect the surface water quality. The first global inventory of maps of Impervious Surface Area (ISA) was produced by NOAA in 2004 (Chand et al., 2009). It covered all man-made surfaces such as parks, parking lots, streets, driveways, sidewalks and buildings. This map used the radiance calibrated DMSP-OLS dataset, street and road density from USGS and land cover classes from 1990s (Vogelman et al., 2001). Later, Elvidge (2007e) produced a global grid of constructed surfaces using radiance calibrated DMSP-OLS dataset, the LandScan data [1](Dobson et al., 2000, Bhaduri et al., 2002) and the 30 m Landsat derived ISA product produced by USGS[2]. The economically developed countries with large areas and total population such as United States, Canada, Norway, Sweden, Finland, Spain, France, Bahrain, Brunei, Qatar and United Arab Emirates exhibited high level of ISA and ISA per person. Japan scored lower in ISA per person because of the topographic and agricultural constraints present in the country (Elvidge et al., 2009a). The country with the most ISA was China (87182 Km^2) followed by United States (83881 Km^2) and India (81221 Km^2). The higher ISA for China and India was attributed to their high population while that of US was for the economic development. In 2010, a new map of impervious surface of South East Asia was prepared using DMSP-OLS imagery and LandScan population data (Sutton et al., 2010).

2.5.2: Application in population studies:

DMSP-OLS night-time images have also been used for population studies. Population studies generally involve estimation of population and population density of an area. Different models such as allometric growth model, regression models, density decay function and TIN (Triangulated Irregular Network) model were used to estimate population from DMSP-OLS images.

Allometric growth model was widely applied to study the relationship between size of a settlement and its population (Doll, 2008). Radiance calibrated DMSP-OLS data was used to estimate the population of China using light area, light volume and percent light as independent variables. Allometric growth models were used to estimate the population at the city and county level as well as the non agricultural population (Lo, 2001a). Regression models were developed from Landsat data and DMSP-OLS data in the 1980s to study the urban population trends in China (Welch, 1980). It provided insights into the Chinese urban policies for cities with a population from 0.5 to 2 million at that time. Later density decay functions and other models were used to study population using DMSP-OLS datasets (Sutton, 1997, Sutton et al., 2001). The night-time stable lights dataset was also extended to estimate the population of every other country of the world (Sutton et al., 2001) based on the original model as proposed by Nordbeck and Tobler (Nordbeck, 1965, Tobler, 1969). The natural log

[1] The LandScan data modelled population globally at 30 arc second (approximately 1 Km) resolution. It modeled people on the basis of their likely position during the day. As a result LandScan used the census data and redistributed them according to the probability coefficients calculated from urban centres, transportation networks, elevation and landcover types (Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. & Worley, B. A. 2000. LandScan: A Global Population database for estimating populations at risk. *Photogrammetric engineering and remote sensing,* 66**,** 849-857.

[2] The 30 m Landsat derived ISA produced by USGS, radiance calibrated DMSP-OLS data and the LandScan data were projected into a 1 Km equal area grid in Albers Projection. Grid counts with population counts of more than 3 were included.

of urban areal extent was correlated with the natural log of population of cities in the study. A further extension of the model was found in 2003 when Sutton described a number of functions such as linear, parabolic, exponential and gaussian functions to explain how population density varied away from the city centres (Sutton, 2003a). Radiance calibrated DMSP-OLS data was used in this study. Radiance values were used as ‘z’ values to create TIN (Triangulated Irregular Network) model for cities of China. Models were made to find the correlation of the selected parameters with the lit area and lit volume of the cities (Lo, 2002). It was found that total population correlated well with the volume of city lights than with area. Model was proposed to estimate population density for both inside and outside the light patches for 2264 counties of China in 1998. Significant correlations were noted between light intensity and population inside the light patch (Zhuo et al., 2009).

The relationship obtained from correlation between city lights and population figures (Elvidge et al., 1997a, Elvidge et al., 1997c) were used in different studies (Doll 1998 and Doll et al 2000 as referred in Doll, 2008). The radiance calibrated DMSP-OLS data was used by Doll and Muller (Doll and Muller, 1999) to estimate population at the country level for United Kingdom and also to investigate population morphology at the city level (Doll et al., 2000).

A project known as the Global Rural Urban Mapping Project (GRUMP) used the night-time lights dataset and population data from the census. (Pozzi et al., 2003, Schneider et al., 2003, Dobson et al., 2000). It aimed to construct a globally consistent database of urban areas by using the night lights with city level census data to estimate population of urban areas (Balk et al., 2005).This dataset was particularly useful for mapping population in large administrative areas.

In addition to estimating population, different products from the DMSP-OLS satellite were also used to propose surrogate census at sub – national levels. High and low fixed gain single orbit DMSP-OLS night-time images were compared to assess their applicability in proposing surrogate census for India (Roychowdhury et al., 2011b). This study provided a relative assessment of gain setting for the DMSP-OLS images in an urban Indian context. Images with a gain of 50 dB proved suitable for larger areas while those with a gain of 20dB produced better results at a smaller spatial scale. Global composites of DMSP-OLS stable lights and radiance calibrated data were also used to propose models for surrogate census at sub national level for India (Roychowdhury et al., 2011a, Roychowdhury et al., 2010). Correlations between stable lights and brightness information with available census metrics from the last Indian census (2001) were calculated using bootstrapping techniques. Linear regression and multivariate analyses were subsequently performed and models proposed for each of the selected census metrics at district and sub-district levels with results ranging from r^2 of 0.4 to 0.9 at the 95% confidence interval.

The variations in light information as obtained from the images were used to study the resource utilization and infrastructure development of an area. Spatial and temporal changes of electric power consumption using DMSP-OLS data was also studied over India during the decade 1992-2003 (Chand et al., 2009). Integrating demographic data with the results from the study it was found that with an increase in population of 170 million during the decade, electric power consumption increased by 26% to more than 300,000 million kilowatt. There was an increase in the light intensities along the peripheries of major cities in the country indicating increased stress on the cities and corresponding development in power consumption pattern over the decade from 1992-2003. The relative merits of GRUMP and LandScan datasets were assessed in a study of areas of the world with no access to electricity (Doll, 2010). Results showed that LandScan data exhibited lower rates of detection of electrification compared to GRUMP data for a given population density. An estimation of steel stock and construction material using DMSP-OLS images was conducted by Hsu et al (2010). As DMSP-OLS data was correlated with human activities, CO_2 emission and electricity use, it was presumed that all these were linked to the amount of steel used in a country. DMSP-OLS data was used in conjunction with land cover data to estimate building and construction steel in Japan, China, South Korea and Taiwan. Building steel stock correlated well with urban night-time lights while the steel used for civil construction produced better correlations with the total light information from DMSP-OLS images. Lit foci obtained from DMSP-OLS night-time data from September 1999 were correlated with population data and electrical power consumption (Amaral et al., 2005, Amaral et al., 2006).

2.5.3: Applications in studying economic activities:

The level of development of an area from its economic activities is recorded by the Gross Domestic Product. DMSP-OLS datasets have been applied in areas of estimating GDP from an area and also for mapping poverty.

The correlation between night-time lights and economic activities was first studied by Elvidge et al (1997c). The light area from DMSP-OLS data was used to plot against Gross Domestic Product (GDP) of a number of countries at different levels of economic development. Maps of Gross Domestic Product – Purchasing Power Parity (GDP-PPP) and Subtotal Ecological-Economic Product (SEP) were prepared using DMSP-OLS images and data on economic activities (Sutton and Costanza, 2002, Ebener et al., 2005, Doll et al., 2000). Doll et al (2006) created maps of coterminous United States and western Europe at 5 Km spatial resolution using night-time radiance data and regional economic productivity data. Since different countries had unique relationships with their light usage based on their cultures, each country was considered separately in this study. The subnational analysis revealed linear relationships between brightness and Gross Regional Product (GRP).Global night-time datasets were specifically used for GDP and Carbon Dioxide (CO_2) emission was correlated with total lit area of a country to prepare the first ever map of GDP-PPP of the world (Doll et al., 2000) It was found that the tropics and wetlands were the hotspots arising from the first map of SEP (Sutton and Costanza, 2002). The effectiveness of DMSP-OLS night-time datasets to estimate GDP was further extended when a new global model for the prediction of income per capita at country level was proposed by Ebener et al (2005). Along with using light as one of the parameters, other information obtained from the DMSP-OLS dataset such as number of cells with light and total frequency of observation were used in the model. It showed that DMSP-OLS data provided an independent measure of economic activity. A close relationship was found between GDP per capita of countries of the world and Information and Communication technology Development Index (IDI) (Ghosh et al., 2010). A disaggregated map of GDP per capita was prepared as a first step using radiance calibrated DMSP-OLS night-time images and LandScan population grid of 2006. As the GDP per capita was less at the bright city cores, there was no correlation between the city cores and the official IDI figures. As a result, IDI maps were created at the state level for south east Asian countries (Ghosh et al., 2010).

In addition to these application, DMSP-OLS dataset has also been used for mapping poverty. A global map of poverty was produced by using the quantity of lighting per person as an indicator of poverty level (Nghiem et al., 2009, Elvidge et al., 2009b). Countries such as India, China and in Africa showed preponderance of poverty according to the map. Ethiopia, Madagascar, Uganda, Tanzania, Malawi, and Niger with a total population of 10 million exhibited more than 80% poverty rates. On the other hand, countries such as Taiwan, South Korea, Netherlands, Italy, United Kingdom, United States, Canada, Chez Republic, Germany, Greece, Spain, Hungary and France had less than 10% of poverty incidence.

2.5.4: Applications in environmental studies and disaster management:

Another area of application of DMSP-OLS dataset was in environmental studies and disaster managements. For example, the radiance information obtained from DMSP-OLS images were used as a surrogate to map and monitor greenhouse gas emissions such as carbon dioxide (CO_2) and Nitrogen Oxide (NO_x) emissions.

Elvidge (1997c) first used DMSP-OLS data to monitor carbon dioxide emission. Later Doll (2000) used this dataset to map carbon dioxide emission at a global scale at 1^o by 1^o resolution. The map was compared with the carbon dioxide emission map produced by Carbon Dioxide Information Analysis Centre (CDIAC). The global carbon dioxide map produced by Doll (2000) from night-time lights showed smaller range of emissions and estimated the emission of CO_2 25% less than that of CDIAC map in areas such as eastern United States, Europe, India and China. A number of outliers were identified in the study. While Doll et al (2000) hypothesized that these areas might have lower than average levels of street lighting and many energy inefficient industrial activities (Doll, 2008), Saxon (1997) noted that the outliers were small islands which had extreme concentration of carbon dioxide sources. As a result the emissions could not be related to the illumination. Saxon (1997) used the correlation of lights and CO_2 emission as a model to track country compliance with the Kyoto protocol. Any change in the emission could be tracked over short period of times by observation of a country's departure from the predicted values along with the trend in carbon dioxide emissions reported by the country's official figures.

The stable lights dataset along with three independent data sources such as the measurements from Global Ozone Monitoring Experiments (GOME), global 3D Chemical Transport Model (CTM) and Emission Database for Global Atmospheric Research were used to study NO_x emissions (Toenges-Schuller et al., 2006). Nitrogen Oxide distribution was taken as a surrogate of the distribution of corresponding anthropogenic emissions from fossil fuels and industrial wastes. The light data was used as a proxy for anthropogenic nitrogen oxides emission with the assumption that many people were living in lit areas and emitting substantial amount of nitrogen oxide from industrial infrastructure.

Along with studying the greenhouse gas emissions, DMSP-OLS night-time images were also used to study the effect of light on animal species from urban and other ecosystems (Cinzano et al., 2001). Ecological light pollution included chronic or periodically increased illumination or unexpected change in illumination or direct glare (Longcore and Rich, 2004). The amount of light pollution was first mapped by Cinzano et al (Cinzano et al., 2001) who used radiance calibrated night-time images to quantify artificial night sky brightness. Upward flux was calculated from the radiance obtained from DMSP-OLS images. The propagation of light pollution was modeled using Raleigh scattering by molecules, Mie scattering by aerosols, atmospheric extinction along light paths and Earth curvature. It was found that majority of the world population, especially in parts of United States and Western Europe, lived in areas above the threshold limit of pollution. Light illumination affected patterns of community ecology by affecting species interactions such as frogs, moths, birds, beetles, flow worms and bat (Longcore and Rich, 2004). Many bats were attracted to insects congregated around lights while many high flying birds tended to lose their way and encounter mortality colliding with building windows when they flew low at night. However, one of the species which was mostly affected by changes in light illumination was sea turtles. The hatchling sea turtles were disoriented because of the presence of lights on sea beaches along with their predation (Salmon, 2005, Salmon et al., 2000). Impact of light pollution on other species such as seabirds, reptiles, zooplanktons and other amphibians were also noted (Perry et al., 2008, Tuxbury and Salmon, 2005, Wiese et al., 2001, Montevecchi, 2006).

The powerful lights fishermen used to attract squids to surface water helped in detecting shipping fleets by the DMSP-OLS images. The fishing of jumbo flying squids (*Dosidicus gigas*) was quantified using DMSP-OLS night-time lights off the coast of Peru in eastern Pacific Ocean (Waluda et al., 2004). A methodology to monitor fishing boats from composite DMSP-OLS image was proposed by Nagatani (2010). A semi real-time DMSP-OLS night-time data product was generated operationally from the Satellite Image Data Base system (SIDaB) by Agriculture, Forestry and Fisheries Information Technology Centre (AFFRIT), Japan. This data was used to propose a method to prepare DMSP-OLS mosaic images with fishing boats. DMSP-OLS image was able to estimate the number of vessels on Peruvian water in 83% of the cases. This study offered an important measure to police shipping vehicles in the Exclusive Economic Zone (EEZ) in that zone.

DMSP-OLS data was used to study the impact of city on soil resources. The urban footprints obtained from DMSP-OLS data along with census data and census soil maps was used to plot the extent of urban built up area and its potential impact on the soil resources in conterminous United States (Imhoff et al., 1997a). City lights were mapped from DMSP-OLS datasets in non-agricultural areas where *population density* was 947 persons per square kilometre and housing density of 392 housing units per square kilometre. Comparing with the Food and Agriculture Organization of the United Nations (UN/FAO) Fertility Capability Classification System, it was found that the most fertile soils were under urban areas and some unique soil types were under the threat of loss because of urbanization.

In disaster management DMSP-OLS data was used especially in the areas of detecting forest fires. Other potential areas included detecting the loss of street lights, buildings or power stations resulting in lower levels of light being detected after any natural disaster striking a region (Doll, 2008). Forest fires could be detected in DMSP-OLS datasets from their non-urban location and ephemeral nature (Elvidge et al., 1997b). The detection of forest fires using DMSP-OLS data was first proposed by Elvidge et al (2001a). Calibrating DMSP-OLS data showed that 77% of the heavily lit forest area in Roraima region was outside the areal limit of the protected forest in those areas. Chand et al (2006) used DMSP-OLS data to detect forest fires over India. India had 55% of total forest prone to forest fire. The results obtained from DMSP-OLS data was compared with the ground observations of fire records from fire departments. Comparing with MODIS and AWiFS sensor on board the Indian Remote Sensing satellite (IRS-P6) the results from DMSP-OLS showed 98% accuracy. Chand (2006) proposed an algorithm for fire detection in identifying pixels which lay outside the stable light datasets from every individual pass of DMSP-OLS datasets. Later fires were detected in the state of Kerala in

India using DMSP-OLS dataset along with fine –resolution Indian Remote Sensing (IRS) – P6 Advanced Wide Field Sensor (AWiFS) and MODIS derived fire product (Badarinath et al., 2011). Results showed a strong correlation between DMSP-OLS derived fire counts and MODIS fire product. The Early Damage Area Estimation System (EDES) was developed by Kohiyama et al (2004). This was currently the only known real time use of DMSP-OLS images for disaster management. The study (Kohiyama et al., 2004) detected loss or reduction of lights from cities affected by natural disaster. It used either two or time series of images to estimate global urban damage and provided geographical information through the internet within 24 hours after a severe disaster event.

2.5.5: Atlas of night-time sky:

An important application of the night-time datasets was to prepare atlas of night sky brightness. Attempt was made to prepare the first census of night sky brightness in 2001 (Cinzano et al., 2001). A second atlas of night-time sky brightness (Cinzano et al., 2007) altered all the major steps from the first atlas. For example, extended Garstang models were used to provide simple solutions to the problems of radiation due to atmospheric pollution. Radiance from DMSP-OLS images was measured using a new software package called LPTRAN which was capable of recording radiation flux and scattered flux in a 3D grid. The number of visible stars in a clear sky was added as an input indicator for measurement. A new dataset from DMSP-OLS was produced for this study. This new dataset was composed of geolocated images from visible band in 30 arc second grids with both low gain and high gain which extended to view large angles at about 80 metres from the nadir, scan angles of the OLS sensor towards the earth, times of observations of each scan line, thermal band images, brightness temperature difference in the thermal band images, temperature models, lunar illuminance and the labelling of snow and clouds on the data. In order to know the upward emission functions from the cities, three components were taken into considerations. They were the satellite measurements of the standard sources of lighting on the earth surface with known photometry, earth-based measurements of the upward emission function from surfaces and also modelling of cities by summing the upward emission functions from the lighting sources randomly installed. Lambertian sites lighted by moon were used as reference for obtaining the standard sources of lighting on the earth surface. An extension of the software Road pollution, used for the evaluation of environmental impact of light pollution, was used for modelling of cities.

However, most of the applications of the DMSP-OLS datasets looked into larger spatial scales. This research explores the utility of the night-time images in proposing census for small spatial scales. As a result, the study looks into administrative regions at sub national level. Also, most of the previous researches using did not look into detail in the effects of MAUP and scale effects. This research explores the data quality issues and errors that arise from using different sets of spatial data.

2.6: Limitations of DMSP-OLS Datasets:

Although the DMSP-OLS datasets have been used in different applications across the globe, this data suffers from a number of shortcomings (Elvidge et al., 2007c). One of the most important shortcomings of the dataset is the coarse spatial resolution. The “fine” mode of DMSP-OLS data with a spatial resolution of 0.56 Km is also called the full resolution data. There is an on-board averaging of five by five blocks of fine mode data to produce the “smooth” data with a Ground Sampling Distance of 2.8 Km. The DMSP-OLS sensor lacks onboard calibration. There is absence of systematic recording of in-flight gain changes. The stable light dataset and the average DN dataset as obtained from the DMSP-OLS sensor have 6-bit quantization. Only the radiance calibrated dataset (Elvidge et al., 1999) has 8 bit quantization. The areas around big urban centres have pixel saturation, commonly known as the blooming effect (Imhoff et al., 1997b). This tends to increase the lit area more than the actual urban area. There is a lack of thermal band to help fire detection. The data also suffers from limited dynamic range, lack of well characterized Point Spread Function (PSF) and Field of View (FOV). The data is available in single band. Absence of multiple bands prevents differentiation among the various sources of lights used. Most OLS data are averaged on board to facilitate download of global coverage. Hence there are limited options of data recording and downloading (Elvidge et al., 2007b, Elvidge et al., 2007c).

In order to overcome these shortcomings, the Visible/Infrared Imager/Radiometer Suite (VIIRS) has been proposed to fly on the National Polar-orbiting Operational Environmental Satellite System (NPOESS). The VIIRS is a 22 channel imager. It has a Day/Night Band (DNB) which is a visible channel capable of imaging the earth and its atmosphere in both bright solar illumination and nocturnal lunar illumination. Similar to the DMSP-OLS, this DNB is capable of detecting clouds at night, image patterns of urban development by recording emissions from cities. It can also record emissions from fire along with snow and ice on the surface of the earth (Lee et al., 2006, Miller et al., 2005, Miller and Turner, 2009, Schueler et al., 2002). VIIRS provides low noise, calibration and stability, broad spectral coverage, wider dynamic range, higher quantization, on-board calibration and fine spatial resolution (0.742 Km). In addition to improved radiometric, spectral, and spatial performance, VIIRS features global coverage twice-daily in each orbit plane (Schueler et al., 2002, Elvidge et al., 2007c). Although the VIIRS is an improvement over the DMSP-OLS, it is imperfect to measure nocturnal lighting from human settlements as the resolution of 0.742 is coarse for identifying different features within human settlements (Elvidge et al., 2007c, Lee et al., 2006).

2.6.1: Other sources of night-time datasets:

There has been attempts to capture images at night from sources other than DMSP-OLS satellites. Photos taken from International Space Station (ISS) are the only other space borne source of night-time images. During expedition 6 (23 November 2002 – 3 May 2003) to the ISS, a manually driven “image motion compensation mount” was built and photos of cities were captured at night through “an optical-quality nadir-viewing window” (Elvidge et al., 2007b). An IMAX movie camera mounted with both “a 70 mm film format (55 mm square image) Haselblad 203 camera with 350 mm telephoto lens (a square field of view of 9 degrees) and a Kodak-Nikon 760 digital camera with 58 mm, f1.2 nocto aspheric lens was use to take the images (Elvidge et al., 2007b). Although all crews on board the ISS attempted to take photos at night, it was only Don Petit who built this tracking equipment and was able to capture the “crispest images” from ISS (Doll, 2008). Samples of these images can be viewed at NASA website (NASA, 2009) and details of ISS missions and astronaut photography can be obtained from the website of Johnston Space Research Centre in Houston, US (Image Science and Analysis Laboratory, 2009). There are some drawbacks of astronaut imagery such as absence of precise geolocation information beyond half degree decimal precision around the centre half of the co-ordinates, manual capture of the images, lack of consistent spatial resolution (estimated to be 6m/pixel) and qualitative interpretation of brightness differences (Doll, 2008). However when compared with DMSP-OLS images, astronaut photographs provides more details such as the alignment of road networks and the major highways. The night-time imagery taken over Denver from the ISS has been used to determine the correlations between population and *population density* over Colorado (Anderson et al., 2010). However, it was found that the correlations were stronger with DMSP-OLS images than those with the finer resolution space photographs. The importance of these photographs lay in studying other complex phenomena such as ambient *population density* and impervious surface.

In the absence of other space borne sources, attempts have been made to mount sensors on airplanes flying on high altitudes. The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is a sensor that was capable of recording high resolution images of cities at night. The AVIRIS has 224 very narrow bands (approximately 10 nm) from 0.41 – 2.45 um (Porter and Enmark, 1987). This sensor is designed to fly on board the NASA’s U2 aircraft and is capable of imaging at a pixel resolution of 20 m over a 10 Km swath at an altitude of 20 Km from the earth surface (Doll, 2008). The AVIRIS is capable of detecting differences in spectral signatures over cities such as Las Vegas where a test flight was conducted in 1998 (Doll, 2003, Elvidge, 1999, Doll, 2008). The commonly used light sources such as high pressure sodium for street lighting, mercury vapour and halide lamps for sports stadiums and car parks had different spectral properties. Mapping of spectral patterns over cities enabled differentiating between the various landuses such as residential, commercial, industrial and recreational including parks and open spaces (Doll, 2008, Elvidge, 1999). Imaging spectrometer data was used to identify and characterize urban lights in Las Vegas (Kruse and Elvidge, 2011). ProSpecTIR imaging spectrometer data was used to extract spectral features from different sources of lights. The results were compared with the spectral signatures of DMSP-OLS datasets on a pixel by pixel basis and a map of nature and distribution of types of lights was prepared.

Based on high resolution field spectra of outdoor lighting and moderate resolution space photographs of the earth at night taken from the ISS, a satellite system capable of global observation of the primary features of human settlement has been proposed (Elvidge et al., 2007c, Elvidge et al., 2007d, Elvidge et al., 2008, Elvidge et al., 2007b). This is known as the Nightsat Mission. A spatial resolution of 50 to 60 m has been suggested for the Nightsat Mission. In this mission, images from Cirrus with a spatial resolution of 1.5m were used to simulate images with 25, 50, 100, 200 and 742 m Overall spatial resolution of 50 to 60 m was found to be optimum to map urban forms and detect the annual development growth rates (Elvidge et al., 2007b). At this spatial resolution and a swath of 80 – 90 Km, there would be a revisit period of 29 days matching the synodic period of the moon at the equator (Elvidge et al., 2007b, Doll, 2008). The overpass time would be 9:30 pm and like DMSP-OLS datasets, cloud and fire screening would be done using a separate thermal band (Doll, 2008). Three types of outdoor lighting sources were identified such as flames produced by combustion, incandescent light produced by heated filaments and vapour lamps producing light from electrically charged gases. Most commonly used lightings around the globe include incandescent, mercury vapour lamps, high pressure sodium vapour lamps, low pressure sodium vapour lamps, metal halide lamps and Light-Emitting Diodes (LEDs). All these light sources had distinct spectral signatures. For the nightsat mission a broad low light imaging band with a bandpass from 0.5 – 0.9 μm (commonly referred to as panchromatic band) was proposed. This bandpass was similar to the low light imaging bands used in the DMSP-OLS sensor. The radiance detection limit for nightsat mission was proposed to be $2.5E^{-8}$ watts/cm^2/sr/μm to a saturation radiance of $2.5E^{-2}$ watts/cm^2/sr/μm. A thermal band with a bandpass from 3 to 5 um might be incorporated in the nightsat mission if the sensor was flown separately and not on a multi sensor platform. This bandpass range was recommended after examining MODIS Airborne Simulator Data over Los Angeles, California (Elvidge et al., 2007b). The nightsat performance metrics as proposed by Elvidge et al (2007b) is shown in table 2.7.

Table 2.7: Different specifications of Nightsat Mission (Elvidge et al., 2007b p. 2665)

	Low-light band(s)	Thermal band(s)
Ground sample distance (m)	50–100	300–500
GIFOV (m)	70–150	400–600
Swath (km)	~200	~200
Revisit time (days)	~20	~20
Geolocation error (RMS in m)	50	100
Min. radiance or temperature (Watts $cm^{-2} sr^{-1} \mu m^{-1}$ or degrees K)	Good: $2.5E^{-8}$ (human settlements) Better: $2.5E^{-9}$ (terrain lit by shielded lighting and sparsely lit development) Best: $5E^{-10}$ (artificial sky brightness)	240 K
Max. radiance or temperature (Watts $cm^{-2} sr^{-1} \mu m^{-1}$ or degrees K)	$2.5E^{-2}$ (sunlit terrain)	400 K
Duty cycle	40%	40%

2.7: Summary:

Remotely sensed datasets have been used widely for identifying and characterizing human settlements and associated socio-economic metrics. However, many research challenges remain. The criticisms of using remotely sensed images to estimate population was put forward by Morrow-Jones and Watkins. They pointed out that collection of higher resolution aerial images over a country as big as US was very time consuming. Another limitation pointed out was that population data obtained from remotely sensed images was neither based on the place of legal residence nor on the place of the location of a person at the time of counting. This restricted the comparison of the population figures with past published census figures. Another important drawback of using satellite images for predicting census was the lack of accuracy. Lo (2006) proposed that if a census area was divided into fairly small parts, then the estimated population for each small area would add up to quite an accurate population estimated over the whole census tract. But, the most important criticism as put forward by Morrow-Jones and Watkins (1984) was that other characteristics of population such as age, caste, literacy, working population, migration and household income, as provided by the census, could not be obtained from remotely sensed datasets. These useful data needed to be supplemented with field surveys. On the basis of this point, it was proposed that remote sensing should be used as a complementary tool along with the traditional method of census estimation because it has an important role in improving the accuracy of traditional censuses.

The DMSP-OLS dataset is presently the only dataset available that records light emitted from cities at night. The data proved to be suitable in studying growth of urban area, economic development, environmental conditions, species behaviour and disaster management. This research explores the utility of radiance calibrated and stable light DMSP-OLS images to propose the census at three different spatial scales. The areas such as utility of the dataset at sub national levels and MAUP are looked in details in this research. The details of the data processing, results and application of the models are described in the following part of the thesis (chapter 4 to chapter 7).

3. Study Area

The research is focused on the state of Maharashtra in western India as a case study. Maharashtra is the third largest state in India with an area of 307,713 square kilometre (India Planning Commission., 2007). The state is located between 16° N and 22° N latitude and 72° E and 80° E longitude. It is surrounded by the states of Gujarat and Madhya Pradesh in the north, Chhattisgarh in the east and Andhra Pradesh and Karnataka to its south (figure 3.1). Arabian Sea lies to the west of the state, which has a 720 Km long coastline stretching from Daman in the north to Goa in the south (figure 3.1). This is the second most populated state with 9.42 percent of the total population of the country living there (India Planning Commission., 2007). The state exhibits a range of urbanization and population distribution. It is the second largest urbanized state of India with 43 percent of its people living in urban areas (India Planning Commission., 2007). The capital of the state is Mumbai which is the largest urban conurbation of the country. It is the principal financial centre and a major commercial hub of India.

3.1: Maharashtra: A Geographical Profile:

The state of Maharashtra is topographically divided into five broad regional groups or socio – cultural units (India Planning Commission., 2007). These are: Greater Mumbai, Western Maharashtra, Marathwada, Konkan and Vidarbha (figure 3.1). The headquarters of each division are Navi Mumbai, Nashik, Pune, Aurangabad, Nagpur and Amravati respectively. The districts of Mumbai, Thane, Raigad, Ratnagiri and Sindhudurg consist of the Konkan division (along the coast shown in dark yellow in figure 3.1). This region is characterized by small land holdings and no significant irrigation facilities (India Planning Commission., 2007). Similar land holdings are also found in the Pune sub division comprising if the districts of Pune, Sangli, Satara, Kolhapur and Sholapur districts (shown in pink colour in figure 3.1). These districts lie in the rain shadow area and have mainly canal and well irrigation facilities (India Planning Commission., 2007). In contrast, the Nashik sub division is characterized by large land holdings, forests, few fertile lands and good rainfall (shown in dark green in figure 3.1). This sub division is mainly occupied by a tribal population and consists of the districts of Nashik, Dhule, Nandurbar, Jalgaon and Ahmadnagar (India Planning Commission., 2007). The districts of Aurangabad, Jalna, Parbhani, Hingoli, Nanded, Osmanabad, Bid and Latur comprise the Marathwada sub division (marked in light yellow in figure 3.1). These districts have rocky and dry soils with very little rainfall (India Planning Commission., 2007). The Vidarbha subdivision is divided into two parts. One part consists of the districts of Buldana, Akola, Amravati, Washim and Yavatmal and has its head quarters at Amravati (shown in orange in figure 3.1). The second part of the sub division consists of the districts of Nagpur, Wardha, Bhandara, Gondiya and Gadchiroli (districts coloured in light green in figure 3.1) and is under the administration of Nagpur head quarter. Both these subdivisions have moderate to large land holdings, plenty of rainfall and deep black soils. (India Planning Commission., 2007). Bhandara, Gondiya, Gadchiroli and Chandrapur districts have large parts under forest cover and consist of tribal population (India Planning Commission., 2007).

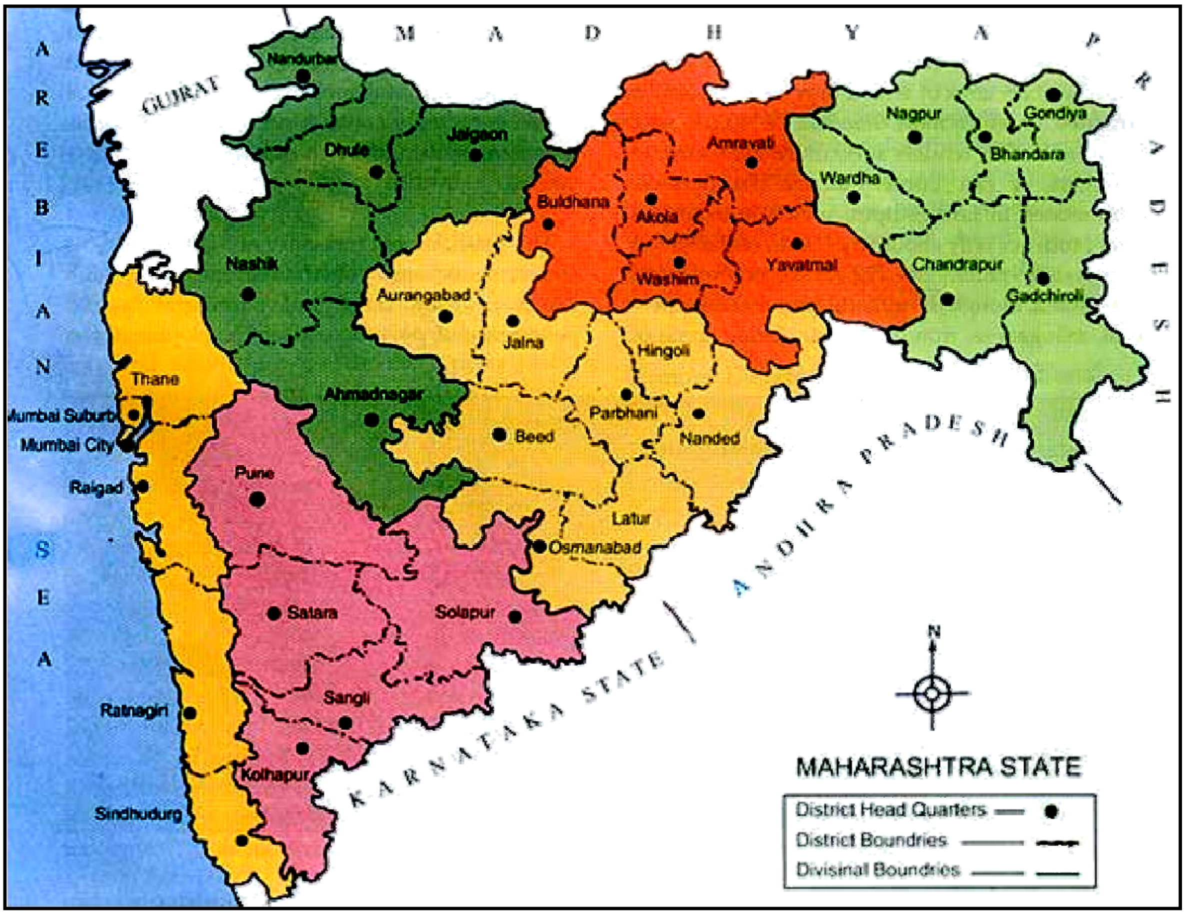


Figure 3.1: Surrounding states and regional sub- divisions of Maharashtra (India Planning Commission., 2007)

3.2: Maharashtra in the Indian Census:

According to the Indian census, the state of Maharashtra is divided into 35 districts with areas ranging from more than 17,000 square kilometres (district of Ahmadnagar) to 157 square kilometres (district of Mumbai) (Office of the Registrar General and Census Commisioner of India, 2001a). The districts are further subdivided into 353 taluks and more than 40,000 villages. Maps showing the areal extents of these regions were provided by the Space Application Centre (SAC), India. The hierarchy of these administrative regions is shown in figure 3.2. Figure 3.2 (a) shows the district divisions of the state. The taluks of one of the districts (the district of Chandrapur in this case) are illustrated in figure 3.2 (b) while figure 3.2 (c) shows the villages of taluk Rajura in the district of Chandrapur.

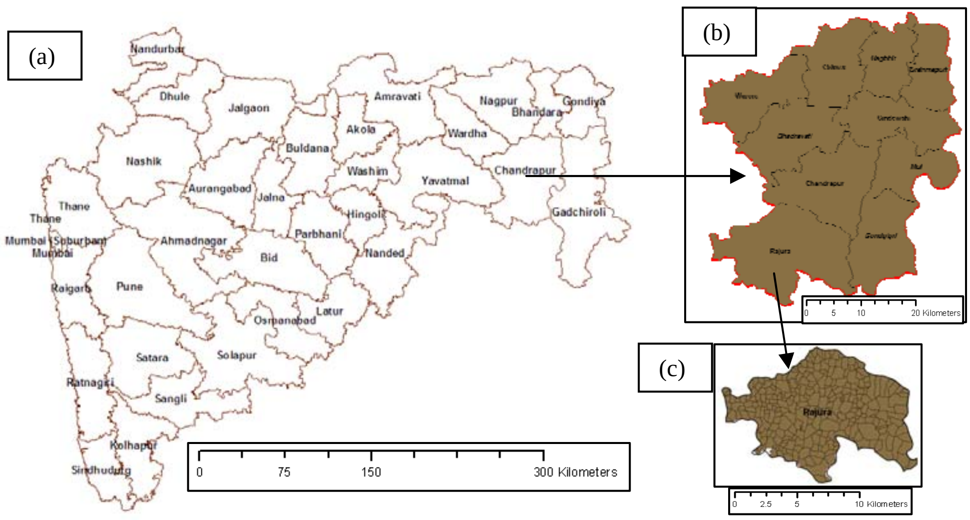


Figure 3.2: Hierarchy of administrative regions of Maharashtra showing (a) Districts; (b) Taluks and (c) Villages

There are more than 19 million households in Maharashtra inhabited by more than 96 million people. Of these 58% of the houses are permanent in nature. There is a 77% literacy rate in the state with 48% of its population engaged in different working categories according to the Indian census. More than 75% of the households in the state use electricity as a power source. However, only 5.8% of the houses have access to their own means of transport such as cars, jeeps and vans. On the basis of the datasets available from the census, an area profile for the state of Maharashtra is presented in table 3.1. Details of the demographic and amenities datasets of Maharashtra as available from the census are presented in Appendix (tables 10.9 and 10.10 to 10.17).

Table 3.1: Selected Census metrics of Maharashtra

Demographic Profile	Number of households	19,576,736
	Total population	96,878,627
	Literates	63,965,943
	Total Workers	41,173,351
	Percentage of urban population	42.4
	Gender ratio	922
	Literacy rate (%)	77.0
Household Profile	Number of factories	296,706
	Number of schools and colleges	178,118
	Percentage of households with electricity as lighting source	77.5
	Percentage of households with telephone	14.1
	Percentage of households with television	44.1
	Percentage of permanent houses	57.8
	Percentage of households with cars, jeeps and vans	5.8
	Percentage of households using banking facilities	48

3.3: Summary

Being the second largest urbanized state in India, there are many big urban areas in Maharashtra. Mumbai Metropolitan Area is the largest conurbation of the country. In addition to this, there are many large cities such as Pune, Nagpur, Nashik and Aurangabad. There are also some very rural areas in the state located in the district of Gadchiroli. The diversity of the state of Maharashtra was also noted in the variation of lights captured in DMSP-OLS images. All these together contributed to making the state of Maharashtra an ideal study area for this research.

4. Preparation of datasets for further analyses

[The method described in this chapter is already published as the following peer-reviewed publication: *Roychowdhury, K, Jones, S, Arrowsmith, C, Reinke, K & Bedford, A 2010, 'Estimating census metrics at a sub-national level using radiance calibrated DMSP-OLS night-time images', paper presented to Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.*]

4.1: Introduction:

This chapter describes the pre-processing of the census and satellite datasets used in this research and the various data quality issues that were encountered. There are four parts in this chapter. The first part illustrates the overview of the methods followed in the study. It is followed by a section on census data processing and associated statistical tests conducted on the available census metrics for the purpose of answering research question 1 (Which socio-economic metrics obtained from the census can be correlated with light information obtained from DMSP-OLS satellite images?). The statistical tests include the process of sampling, tests for normal distribution and bootstrapping the correlation coefficients. The second section describes the satellite image processing where image pre – processing and image analyses steps are discussed in detail. In the third section, the details of the data quality elements are described. The final section (section 4.8) of this chapter combines the results from sections 4.5 and 4.6. Final short listing of the ten census metrics are explained in part four. These ten metrics were used for proposing surrogate census in the following chapters.

4.2: Method overview:

The overall method followed in this research is illustrated in figure 4.1. Details of each step and results are explained later in the chapter (section 4.2 to 4.8). The method that was followed for the study can be broadly divided into two parts. The first part deals with the processing of the satellite images and the second part details the processing of the census datasets. Census data had a number of data quality issues which were successfully dealt with. Mean and standard deviation of both stable lights and brightness were used to predict the models.

4.3: Census data Processing:

Socio-demographic data of Maharashtra was obtained from the primary census abstract of census of India. Data from last complete Indian census (2001) was used in this research. The census data pre-processing and results are described in this section.

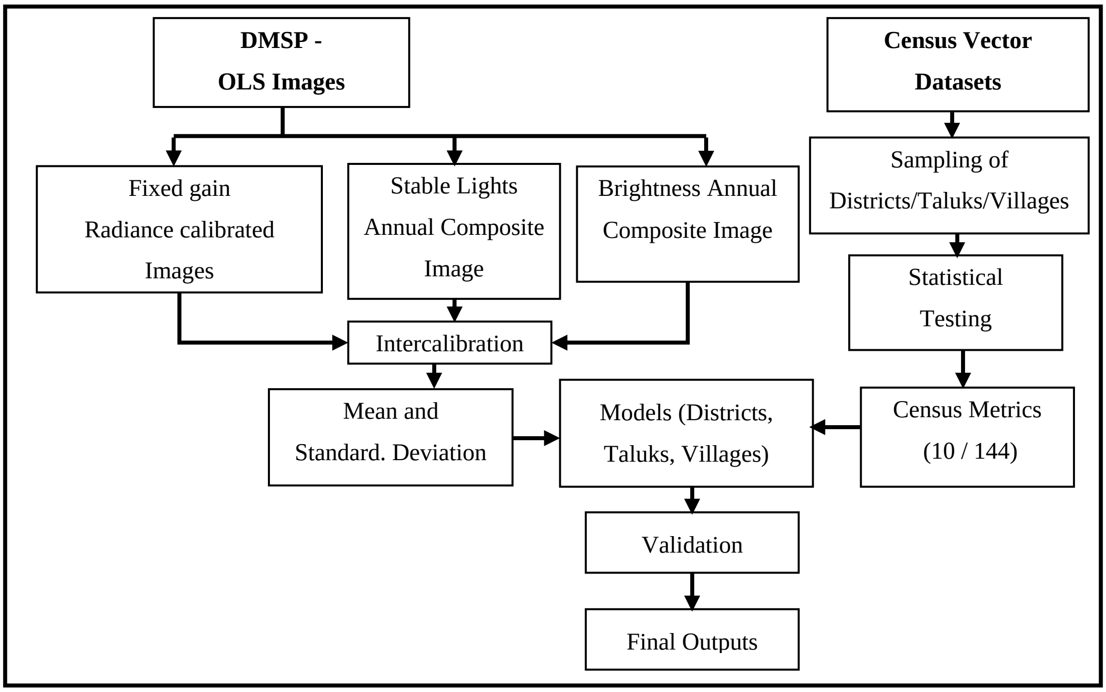


Figure 4.1: Overview of the method followed in the study

Both demographic and amenities data were available from the Indian census for the state of Maharashtra. The demographic data included number of households, total population (both male and female), gender ratio, population composition such as urban and rural population, scheduled castes and scheduled tribes, working population and its sub categories such as main workers[3], marginal workers[4] and non-workers, number of literates and literacy rate (Office of the Registrar General and Census Commisioner of India, 2001a). The amenities datasets accounted for banking facilities, means of transport such as cars, jeeps, vans and bicycles; sewerage system; source of lighting and fuel for cooking; source of drinking water and presence of factories in the area. It also reported on the proportion of occupied and permanent houses in the area, number of rooms in each house, and the roof, wall and floor materials of individual houses.

In addition to the census metrics, an economic indicator, *Per Capita District Domestic Product* (*PCDDP*), was also used in this research. *PCDDP* refers to district level Gross Domestic Product (GDP) (Indicus Analytics Private Limited, 2009). Difference in *PCDDP* between the districts is an important indicator of regional disparity. The Central Statistical Office, Government of India, uses the

[3] Main workers are defined as those workers who work for 6 months or more in a year.

[4] Marginal workers are defined as those workers who work for less than 6 months in a year.

“income originating approach” to calculate District Domestic Product (DDP) that is all the “unduplicated” goods and services produced within a district boundary are taken into consideration (Central Statistical Office, 2009) for calculating the DDP. The *PCDDP* figures at 1998 – 99 price (INR) for all the districts of Maharashtra were obtained from the Maharashtra Development Report (India Planning Commission., 2007).

Although all of these metrics are collected for districts, some of them were unavailable for regions at finer spatial scales. At the taluk level only the demographic metrics and some of the amenities data were present. Economic indicators including *PCDDP* were unavailable for taluks. At the village level only the demographic variables were collected from the census. As a result the models proposed in this research were restricted to the datasets available at each level.

Of the 35 districts of Maharashtra; Mumbai, Mumbai (suburban) and Thane are not considered in this research because of their complete urban character. These districts form part of the Mumbai Metropolitan Area (MMA), the largest conurbation of the country. From the remaining 32 districts, 24 were sampled and census data from them were used to propose the models. Details of these derivations are explained in the following sections.

4.4: Sampling:

The process of sample selection is a very common practice in research, whereby the sample is assumed to be representative of the whole Population. A Population is defined as the whole dataset or the entire set of observations for a given variable under consideration. In this case, Population refers to the total of 32 districts. A sample is said to be the representative of a Population if it has the important characteristics of the Population (Healey, 2008). Although commonly used, the method of sampling in research suffers from some drawbacks. Two most common flaws of sample research are a) overlooking of the sample errors and b) response – non response bias during selection (Wunsch and Gades, 1986, Bartlett et al., 2001). This research used datasets obtained from secondary sources (NGDC/NOAA for DMSP-OLS images and ISRO and Census Bureau from India for vector datasets and primary census abstracts respectively) and therefore, the response – non response bias error did not directly affect the process of sampling in this research.

There are four commonly used methods of sampling: random sampling, stratified random sampling, systematic sampling and cluster sampling (Zar, 2010, Cochran, 2009). In order to make the sample the most representative of the Population, the method of random sampling is the most popular. This method ensures that every case in the Population has an equal chance of being selected in the sample. However, for big Population sizes, the process of random sampling becomes cumbersome and the method of systematic sampling is preferred. In systematic sampling, the first case is randomly selected and thereafter a systematic method of selection is followed. For example, for drawing a Population of 24 districts from a sample of 32 districts, in systematic random sampling, the first case will be selected randomly and then every 32/24th or 1.33 or 2nd case may be selected. However, the result of systematic sampling does not always represent the Population. In stratified sampling method, the Population is first divided into a number of groups based on some particular traits believed to be common to the group. Then samples are collected randomly from each of the groups. This method ensures that the selected sample is truly representative of the Population. However, this method of sampling requires a complete knowledge of the Population without which it becomes difficult to determine the strata. Lastly, Cluster sampling is a method used to select groups or clusters of cases rather than single cases. The clusters are selected through a number of stages which reduces its chance to be a representative of the whole Population.

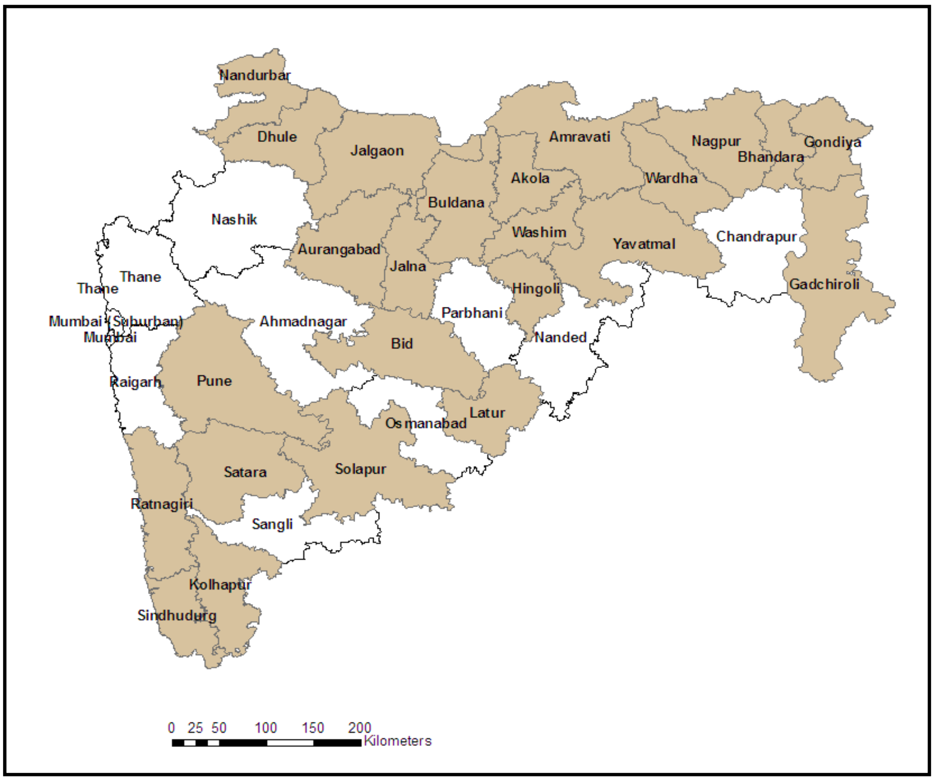


Figure 4.2: Randomly sampled 24 districts of Maharashtra

In this research, the method of random sampling was used to choose 24 districts out of 32 districts. The selected 24 districts are shaded in figure 4.2.

Statistical tests were conducted on the census metrics from these districts in order to shortlist the census metrics that can be associated with the emitted visible light information from DMSP-OLS night-time images. Details of these tests and the results are explained in the following sections.

4.5: Statistical Tests:

4.5.1: Tests for normal distribution

Different methods to test the normality of distribution have been proposed over time. In this study, three kinds of test were undertaken. The first one was to visually examine the histograms and normal probability plots (Zar, 2010), the second test was to examine the skewnwss and kurtosis of the distribution (Field, 2009, Zar, 2010) and the third test of normality conducted was the goodness of fit test using the Shapiro and Wilks statistic (Innes, 2009, Field, 2009).

4.5.1.1: Histogram and Normal Probability Plots:

The method of using histograms and Normal Probability (N-P) plots to test the normal distribution of a variable is referred to as the graphical assessment of normality (Zar, 2010). A histogram is a graphical display of the frequency distribution of the observed dataset in which vertical bars are used to represent the frequencies. In order to test the nature of distribution of a dataset, a frequency curve is superimposed on the histogram. Figure 4.3 demonstrates histograms of 10 variables while the summary statistics for all the metrics are presented in appendix (Table 10.1).

The histograms in figure 4.3 show the variations in the nature of distribution of the different census metrics, the red curve showing what a normal distribution with the same number of data, same mean and standard deviation would look like (Zar, 2010). The histograms show the variation of distribution of the different metrics over the state of Maharashtra. While *total population per square kilometre* and *number of households per square kilometre* show the most normal (bell shaped) distribution, most of the other variables show asymmetrical distribution over the study area. From the histograms it is found that *urban population per square kilometre*, *percentage of households with cars, jeeps and vans, percentage of households with television* and *percentage of permanent houses* have asymmetrical distribution. The distribution of *PCDDP*, as noted from the histograms, was multimodal in nature. A mode refers to the maximum number of frequencies observed in any distribution and is denoted by the tallest bar in a histogram (Field, 2009). A summary of the nature of distribution of the other variables as observed from their histograms is presented in table 10.2 in appendix.

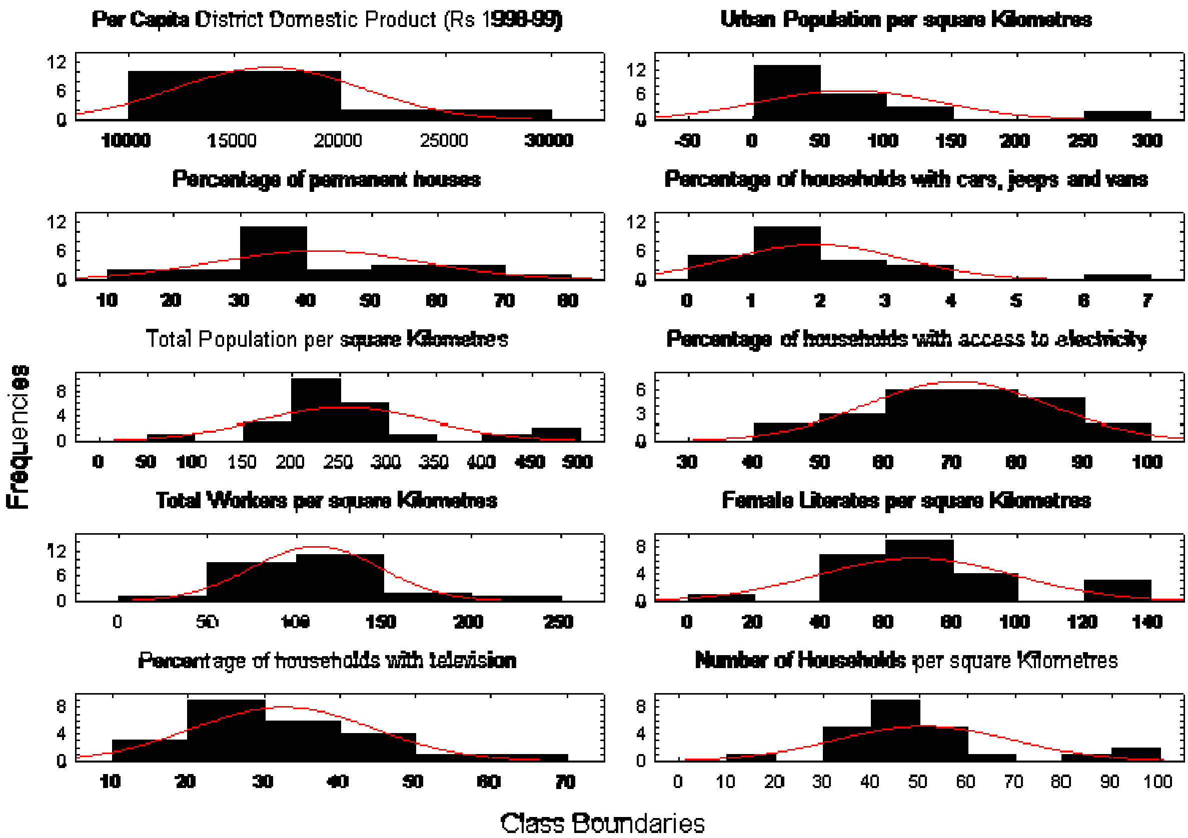


Figure 4.3: Histograms of ten census metrics over the 24 randomly selected districts of Maharashtra

Along with the frequency distribution histogram, the Normal - Probability plot is also an effective measure to determine the normality of distribution. In the Normal - Probability plot, the horizontal axis records the actual data and the vertical axis shows the expected normal value. The normal probability distributions of ten census metrics are shown in figure 4.4. The straight line (shown in red) represents a perfect normal distribution (Zar 2010). None of the variables exhibit much variation from the expected normal distribution. *Percentage of households with access to electricity* and *percentage of permanent houses* is almost normally distributed over the state of Maharashtra. Some outliers are noted in the distribution of the demographic variables such as *number of households per square kilometre, total population per square kilometre, number of female literates per square kilometre, total workers per square kilometre* and *urban population per square kilometre.*

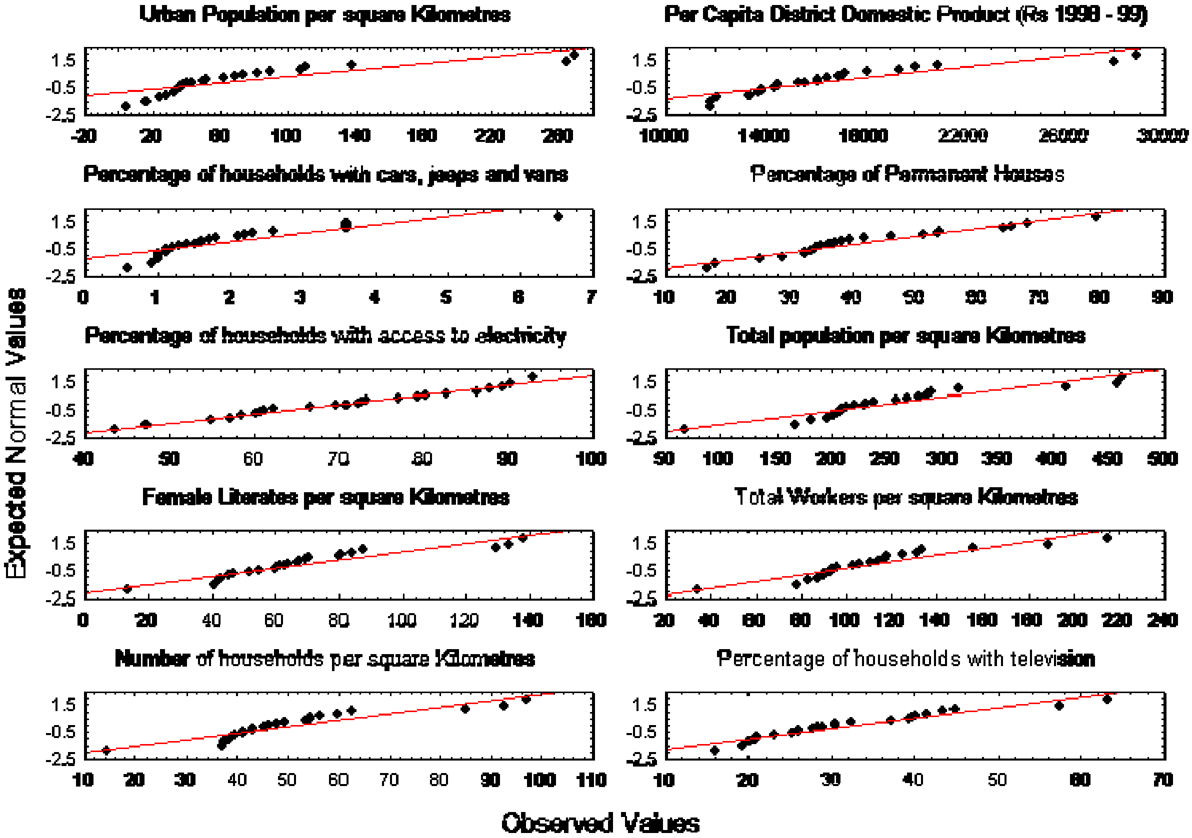


Figure 4.4: Normal - Probability Plots of ten census metrics over the 24 randomly selected districts of Maharashtra

4.5.1.2: Skewness and Kurtosis:

The second test conducted for determining the normal distribution was examining the skewness and kurtosis of the data (Field 2009; Zar 2010). Skewness is a measure of symmetry in the data while kurtosis refers to the peakiness of the data distribution. When data is positively skewed, most of the values are concentrated to the left of the mean and the ‘tail’ of the distribution is to the right. The opposite happens for a negatively skewed distribution with data values concentrated to the right of the mean and the ‘tail’ of the distribution to the left. Positive values of kurtosis indicate a pointed distribution while a negative kurtosis value denotes a flat distribution of data (Field 2009). The values of skewness and kurtosis help to describe the normality of distribution numerically. The further the values are from zero, the distribution is less likely to be normal (Innes, 2009, Field, 2009).

In order to calculate whether the value of zero for both skewness and kurtosis lie within the 95% confidence interval, the z-scores of skewness and kurtosis (Field 2009) were calculated using the following formulae:

$$Z_{skewness} = \frac{S - 0}{SE_{skewness}}$$

$$Z_{kurtosis} = \frac{K - 0}{SE_{kurtosis}}$$

Equation 4.1: Z – scores of skewness and kurtosis

In a perfectly normal distribution the values of skewness and kurtosis are zero. An absolute value of these scores greater than 1.96[5] are significantly different from normal distribution at more than the 95% confidence interval. Therefore, all the variables with z-scores less than 1.96 were considered to be normally distributed. For example, out of the 144 variables considered, 45 variables have confirmed normal distribution according to the results from Z scores of skewness and kurtosis. All of these 45 variables have both the Z-scores less than 1.96 and are shown in table 10.6 in the Appendix.

4.5.1.3: Goodness of Fit: Shapiro and Wilks test:

Another measure of assessment of normality is the goodness of fit test (Zar 2010). The Shapiro and Wilks test were used in this study to test the goodness of fit. This test is used to compare the scores obtained from the sample distribution to a normally distributed set of scores with the same mean and standard deviation (Field 2009). If the results are not significant (i.e. $p > 0.5$), then the distribution is not significantly different from normal distribution (in other words, it is normally distributed). However, if the results are significant ($p<0.5$), it implies that they are significantly different from a normal distribution (or it is not normally distributed) (Field 2009). Another common test used for the goodness of fit is the Kolmorov – Smirnov test. However, for small samples (less than 50) the Shapiro and Wilks test has been preferred by some authors. (Field 2009; Zar 2010). Out of 144 variables tested, 41 are significant at the 95% confidence. The distributions of these 41 variables were confirmed to be normal over the selected sample of districts. The Wilks statistic and the level of significance of these 41 variables are presented in table 10.7 in the appendix.

Twenty nine census metrics conformed to the normal distribution over the study area using the criteria from both the tests. These metrics have the z-scores of both skewness and kurtosis less than 1.96 and significance of Wilks statistic more than 0.05. These 29 census metrics are as follows (table 4.1):

[5] 95% of the area under a normal curve lies within approximately 1.96 standard deviations of the mean. According to the central limit theorem, this number is therefore used in testing approximate 95% confidence intervals.

Table 4.1 : Census metrics with confirmed normal distribution over 24 districts

• Other source of power supply (number of households)	• Percentage of households with Electricity as source of power	• Percentage of female literates
• Female literacy rate	• Gender Gap	• Household size
• Household size per square kilometre	• Number of Industrial Schools per square kilometre	• Percentage of households with Kerosene as source of power
• Main Female Agricultural Labourer	• Main Male Agricultural Labourer	• Main Agricultural Labourer (persons)
• Marginal Female Agricultural Labourer	• Marginal Male Agricultural Labourer	• Marginal Agricultural Labourer (persons)
• Male Marginal Workers	• Total Marginal Workers	• *Percentage of permanent houses*
• Percentage of Scheduled Caste population	• Number of Secondary Schools per square kilometre	• Number of Senior Secondary Schools per square kilometre
• Infant Gender Ratio (0 – 6 yrs)	• Percentage of total female main workers	• Percentage of total male main workers
• Percentage of total main workers	• Percentage of total female workers	• Percentage of total male workers
• Percentage of total workers	• Work Participation Rate	

In addition to these 29 variables, some other census metrics were also chosen on the basis of a number of considerations other than normal distribution. They are as follows:

1. As census is a database detailing information on demographic as well as social variables, some basic demographic variables such as *population density* and *urban population density* are included in this study, although they do not fully conform to the test of normality. These variables are fundamental to any study of the census.

2. Some variables such as *Per Capita District Domestic Product* (*PCDDP*) recorded in Maharashtra Development Report (India Planning Commission 2007) give idea on the economic prosperity of the districts. Although this variable is not normally distributed, it is taken into consideration in this study. DMSP-OLS datasets were successfully used to correlate with GDP per

capita at the global and national level. The variable was included to test the degree of correlation between *PCDDP* and DMSP-OLS derived light information at the sub district level.

3. Variables such a *percentage of households having car, jeep and van* give an idea on the level of development as well as accessibility. This is chosen for analyses although the nature of distribution is not normal. Possession of cars, jeeps and vans is taken as a measure of economic affluence in India. It also provides more mobility to the households to access the centres of main activities such as the nearby towns or urban centres.

In total, 38 variables (shown in table 4.2) were selected and attention would be paid to these variables when conducting statistical comparisons with mean and standard deviation of emitted visible light information obtained from the images. Details of the image processing steps are explained in the next part of this chapter.

Table 4.2 : List of all the census metrics shortlisted for correlation with light information from DMSP-OLS images

• Other source of power supply (number of households)Any Other (num)	• Percentage of households with Electricity as source of power	• Percentage of female literates
• Female literacy rate	• Gender Gap	• Household size
• Household size per square kilometre	• Number of Industrial Schools per square kilometre	• Percentage of households with Kerosene as source of power
• Main Female Agricultural Labourer	• Main Male Agricultural Labourer	• Main Agricultural Labourer (persons)
• Marginal Female Agricultural Labourer	• Marginal Male Agricultural Labourer	• Marginal Agricultural Labourer (persons)
• Male Marginal Workers	• Total Marginal Workers	• *Percentage of permanent houses*
• Percentage of Scheduled Caste population	• Number of Secondary Schools per square kilometre	• Number of Senior Secondary

		Schools per square kilometre
• Infant Gender Ratio (0 – 6 yrs)	• Percentage of total female main workers	• Percentage of total male main workers
• Percentage of total main workers	• Percentage of total female workers	• Percentage of total male workers
• Percentage of total workers	• Work Participation Rate	• *Number of households per square kilometre*
• *Total population per square kilometre*	• *Female literates per square kilometre*	• *Total workers per square kilometre*
• *Percentage of households with cars, jeeps and vans*	• *Percentage of households with television*	• *Percentage of permanent houses*
• *Urban population per square kilometre*	• *Per Capita District Domestic Product* (INR 1998-99)	

4.6: Satellite Image Processing:

The section describes the satellite images used in the study, the image processing and the results obtained from the processed images.

4.6.1: Satellite images used in the research:

Night-time images recorded by the Operational Linescan System (OLS) sensors on board the Defense Meteorological Satellite Programme (DMSP) group of satellites are used in the research. The OLS is a passive remote sensing sensor capable of recording artificial lights from the earth surface at night. This is the only available sensor to date which records night-time lights in the visible and near infra-red (0.4 -1.1 μm) portion of the Electro Magnetic Spectrum (EMS). There are three types of night-time images available from the OLS sensor: the stable light datasets, the radiance calibrated dataset and the average DN product. The details of these images are described in chapter 2 (section 2.1.3). In this research two types of radiance calibrated images (fixed gain single orbit images and global composite) are used along with stable lights images. All of these images are compiled for 2001, the coincident year with the last complete Indian census. The steps undertaken for selecting the images are explained in the following section.

4.6.2: Process of image selection:

I) Single orbit fixed gain radiance calibrated images:

For 2001, 80 single orbit fixed gain radiance calibrated images from satellite F15[6] were provided by NGDC (Tuttle, 2008). The complete list of fixed gain images from satellite F15 for 2001 as obtained from NGDC is available from Table 10.4 in Appendix. Due to the design of the DMSP-OLS sensor, radiance calibrated images are obtained for particular dates only[7] when the satellite is flown with fixed gains.

Using the list of 80 images in conjunction with the Space Physics Interactive Data Resource (SPIDR) archive (http://spidr.ngdc.noaa.gov/spidr/) a list of 12 fixed gain images from satellite F15 were shortlisted. These 12 images were chosen over the study area on cloud free nights. The thermal infra-red images obtained in conjunction with the visible band images were used for cloud flagging (Elvidge et al., 1997a).

Some of the images from the shortlisted 12 suffered from errors such as it was difficult to identify the lighted pixels from low gain images (gain 20). It was, however, possible to differentiate between lit and dark areas from high gain images (gain 50). Also in most cases, the study area was outside the fixed gain portions of the orbits. Although variable gain images (with both fixed and variable gain setting in a single image) for the selected dates were available, there was no method to calculate the radiance from those images. Hence the 'useable' fixed gain images were used in this research.

---

[6] Although satellite F12 and F15 collected fixed gain dataset, data from satellite F15 was only used in this study. DMSP-OLS F12 satellite data had been processed only for US and for the global composite radiance calibrated data for 1996-97 (http://www.ngdc.noaa.gov/dmsp/download_rad_cal_96-97.html). For 2001, only F15 satellite collected fixed gain data some of which were radiance calibrated.

[7] Recently, however, NGDC has prepared monthly composites of DMSP-OLS datasets. These products are not radiance calibrated. To create monthly radiance calibrated product there were three problems: a) It was expensive; b) method of data production was troublesome because of the limited availability of fixed gain data in a month and, c) there was presence of cloud cover regions. This was confirmed through personal communication with the Earth Observation Group (EOG) at National Oceanic and Atmospheric Administration at National Geophysical Data Centre (NGDC/NOAA) Tuttle, B. T. May 2008. *RE: Personal communication.* Type to Roychowdhury, K...

However, due to the limitations of the availability of fixed gain images, only one gain 20 dB image and one gain 50 dB image was used in the study. Both these images were captured over Maharashtra during the month of September, 2001. There were no major differences between these two days in terms of festival, strike or power outages over the study area. The images used in this research were captured on the following dates and times:

Table 4.3 : Dates and times of DMSP-OLS images used in the research

	Gain 20	Gain 50
Date	17^{th} September, 2001	14^{th} September, 2001
Time (GMT)	15:54	14:56

In this research, it was assumed that the radiance information obtained from these two images represented the radiance characteristics for the whole 2001, the census period.

II) Global composite of brightness image of 2001:

The second DMSP-OLS image dataset used in the study is the global composite of brightness data prepared for 2000 – 2001. Fixed gain images taken from satellites F12 to F15 were used to prepare this image. However, this data was not calibrated to radiance (Doll, 2008, Tuttle, 2008) and brightness values for this image ranged from 0 to 653.

III) Selection of stable light image:

The stable light data was obtained from National Geophysical Data Centre (NGDC) website (National Geophysical Data Centre, 2006) using the latest average DN data series. It contains the lights from cities, towns, and other sites with persistent lighting, including gas flares. Ephemeral events, such as fires were discarded from this dataset. Data values in this image range from 1-63, with background noise data replaced with a zero. Areas with zero cloud-free observations are represented by the value 255.

As there was no onboard calibration available on the OLS sensor, the stable light dataset was calibrated using an empirical procedure (Elvidge et al. 2009). It was found that the data collected by F12 satellite had the highest digital values. The data from F121999 was used as the reference data for calibration of datasets captured by other DMSP-OLS satellites such as F14, F15 and F16. Scattergrams were used to locate the areas with the least variation in lighting over time. The graphs over these areas displayed minimal dispersion of lit pixels along the diagonal axis (Elvidge et al. 2009). Images over Sicily presented an ideal case with data spread over a dynamic range and a well defined diagonal axis with less dispersion of light clusters. Figure 4.5 shows the scattergrams of satellite F121999 and F152001 over Sicily.

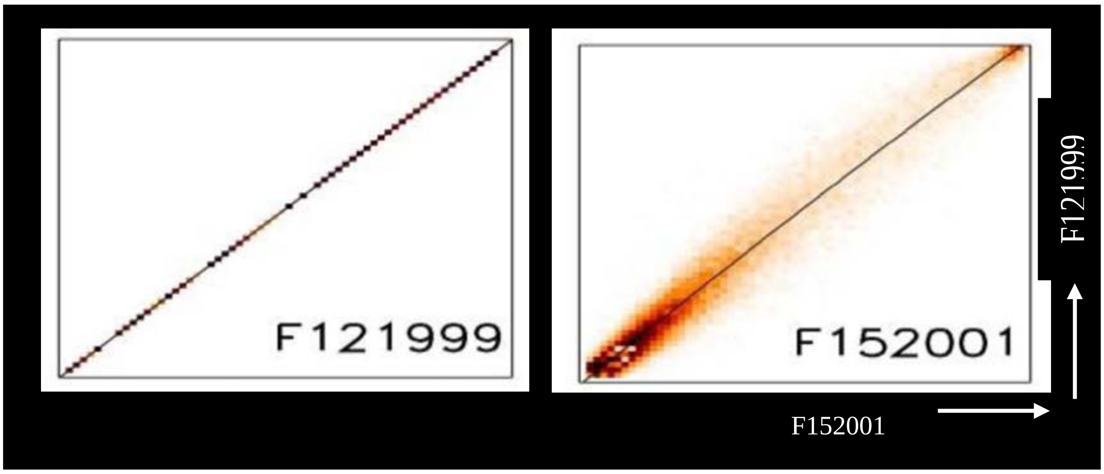


Figure 4.5 Scattergrams from F121999 and F152001 satellites over Sicily (Elvidge et al. 2009 p. 601)

A second order regression model was proposed for intercalibration (Elvidge et al. 2009). For the year 2001, DMSP-OLS night-time images were available from satellites F14 and F15. F142001 and F152001 images were calibrated using the calibration information as obtained from NGDC (Elvidge et al., 2007a). Details of the intercalibration information for all the satellites can be found in Appendix (table 10.3). Mean of the stable light values (after calibration) for every district of Maharashtra is presented in figure 4.6.

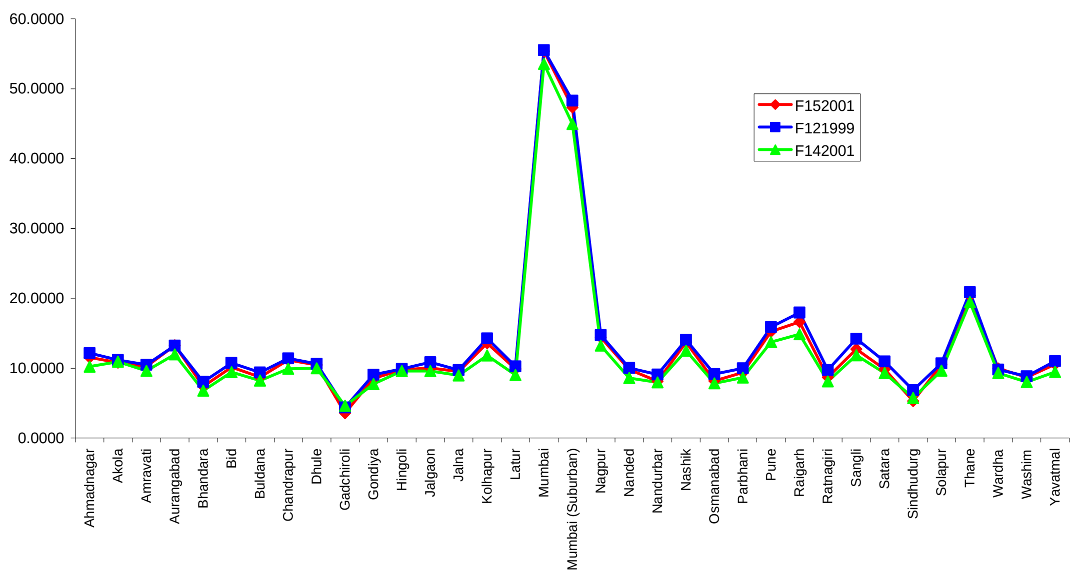


Figure 4.6: Results of intercalibration of stable lights of F142001 and F152001 images

The results from intercalibration are presented in table 4.4. The difference in the mean stable lights between the F14 and F15 images is the highest for Mumbai (suburban) region followed by the districts of Mumbai, Raigarh, Kolhapur, Pune, Ahmadnagar, Aurangabad, Chandrapur, Nashik, Nanded, Nagpur and Yavatmal. Results show that the values for the F15 image were closer to F12 image than those from the F14 image for most of the districts. Image from F14 displayed better results only for the districts of Akola, Gadchiroli and Sindhudurg. Therefore F15 is chosen for further analysis in this study. The shaded cells in table 4.4 indicate the districts with the least difference in the mean stable lights with the F12 image.

After the images were shortlisted, a number of image processing steps were carried out on the images to conveniently extract information from them. These steps of image processing are explained in the next section.

Table 4.4 : Results before and after intercalibration of F142001 and F152001 images. The shaded cells represent the difference in the mean stable lights between two images after intercalibration

DISTRICT	F152001	F142001	F121999	Difference between F15_F12	Difference between F12_F14
Ahmadnagar	11.5204	10.2306	12.1572	0.6368	1.9266
Akola	10.8075	10.9202	11.1724	0.3649	0.2522
Amravati	10.2539	9.6273	10.4939	0.2400	0.8666
Aurangabad	13.2653	11.9799	13.2317	-0.0336	1.2518
Bhandara	7.3941	6.7830	8.0360	0.6419	1.2530
Bid	10.1364	9.4372	10.7519	0.6155	1.3147
Buldana	8.7253	8.2208	9.3872	0.6619	1.1664
Chandrapur	11.1583	9.9186	11.4055	0.2472	1.4869
Dhule	10.4656	9.9981	10.6134	0.1478	0.6153
Gadchiroli	3.5713	4.6243	4.3721	0.8008	-0.2522
Gondiya	8.5487	7.7230	9.0645	0.5158	1.3415
Hingoli	9.7724	9.6030	9.8872	0.1148	0.2842
Jalgaon	10.0351	9.5821	10.8409	0.8058	1.2588
Jalna	9.5211	8.9659	9.7219	0.2008	0.7560
Kolhapur	13.5824	11.8166	14.2576	0.6752	2.4410
Latur	9.8782	9.0268	10.2389	0.3607	1.2121
Mumbai	55.3913	53.5556	55.5263	0.1350	1.9707
Mumbai (Suburban)	47.3394	44.9556	48.2897	0.9503	3.3341
Nagpur	14.4251	13.2277	14.7229	0.2978	1.4952
Nanded	9.8052	8.5990	10.0424	0.2372	1.4434
Nandurbar	8.1297	7.9607	9.1005	0.9708	1.1398
Nashik	13.7210	12.5107	14.0509	0.3299	1.5402
Osmanabad	8.1737	7.8285	9.1423	0.9686	1.3138
Parbhani	9.3949	8.6468	9.9804	0.5855	1.3336
Pune	15.2609	13.7263	15.8962	0.6353	2.1699
Raigarh	16.6277	14.8561	17.9634	1.3357	3.1073
Ratnagiri	8.7127	8.0978	9.7163	1.0036	1.6185

Sangli	12.7327	11.8416	14.2125	1.4798	2.3709
Satara	9.8436	9.3144	10.9485	1.1049	1.6341
Sindhudurg	5.3071	5.7527	6.8509	1.5438	1.0982
Solapur	10.5097	9.6459	10.6940	0.1843	1.0481
Thane	20.3710	19.4596	20.8599	0.4889	1.4003
Wardha	9.9661	9.3179	9.7980	-0.1681	0.4801
Washim	8.7119	8.0227	8.8238	0.1119	0.8011
Yavatmal	10.5519	9.4583	11.0036	0.4517	1.5453

4.6.3: Image Pre - Processing:

In this research, the following image pre-processing steps were conducted:

i. Creating the subset of the study area;

ii. Image geometric correction;

iii. Image Enhancement.

Other image processing such as filtering was not considered suitable for this research as they tended to modify the necessary data values in the images.

i) Creating the subset of the study area:

The DMSP-OLS images are collected on a daily basis and are available from -180 degree to +180 degrees longitude and -65 degree to +65 degree north latitudes. The study area of the state of Maharashtra was extracted as a subset from portions of the orbit covering parts of India. This helped to concentrate only on the area of interest and reduced the amount the data handled for further analyses.

ii) Image geometric Correction:

When obtained from NGDC, the DMSP-OLS images were orthorectified. As a result they were free from errors created due to curvature of the earth, sensor tilt, lens distortion and elevation in topography (Bhatta, 2008). The geometric registration of the images was necessary to use them with other spatially referenced datasets. The DMSP-OLS images were reprojected from World Geographic System (WGS) Projection to Lambert Conformal Conic Projection with the datum WGS 1984. The projection was similar to those of the vector datasets of state and districts of Maharashtra as obtained from Space Application Centre (SAC), ISRO in India. This projection used two standard parallels (12.47 degree north and 35.17 degree north in this case) and one central meridian (80 degree east for this study). The conformal projection maintained the shape and scale of the area projected.

The distorted images were then geometrically corrected by a process of co-ordinate transformation. Image-to-image correction was employed using third order polynomial equation. 58 ground control points were selected from all over the state. The Root Mean Square Error (RMSE) obtained in pixel points was 0.01684 (around 16.84 metre) after the transformation. Generally, a RMSE of between 0.5 and 1 pixel (i.e. 500 to 1000 metre for DMSP-OLS images) width is considered acceptable (Thomlinson et al. 1999, Armston et al. 2002, Reinke and Jones 2006, Wulder et al. 2008). RMSE was calculated using the formula:

$$RMSE = \sqrt{\frac{D_1^{\ 2} + D_2^{\ 2} + D_3^{\ 2} + \ldots\ldots + D_n^{\ 2}}{n}}$$

Equation 4. 2: Root Mean Square Error

Where D_1, D_2, D_3… are the difference in distance between the image (pixel) co-ordinate and the ground co-ordinate and n is the number of points (Bhatta, 2008).

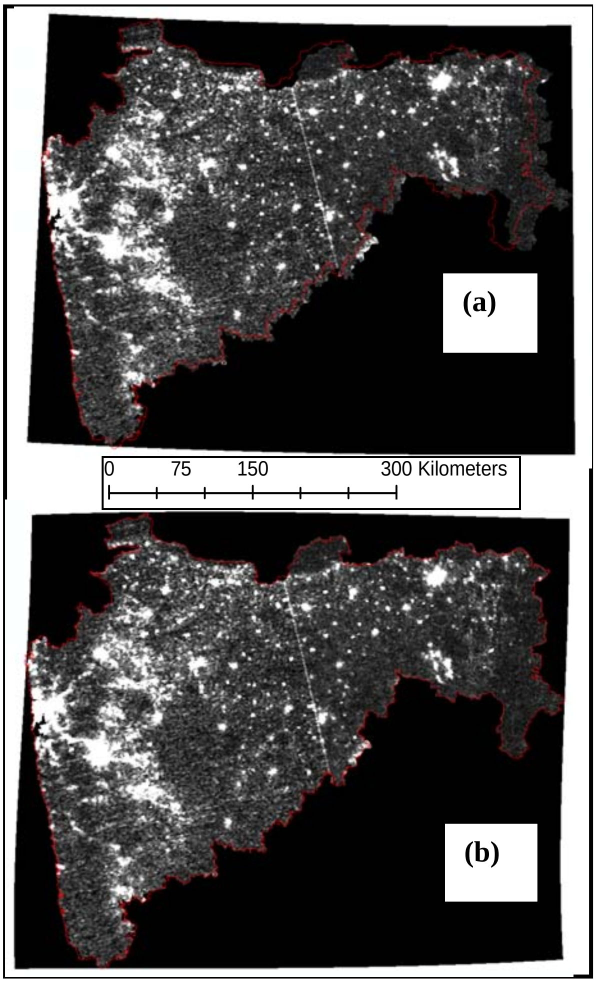


Figure 4.7: Datasets before (a) and after (b) georeferencing

A nearest neighbour resampling method was used to determine pixel values in the output image matrix (Lillesand et al., 2004). This approach used the values of the closest input pixel as the output pixel value. Although this method retained the original values, some pixel values were lost and others duplicated in the output image (Bhatta, 2008). It also gave a "disjointed appearance" to the image (Lillesand et al., 2004).

The annual composite image needed some editing before further analyses. In the original annual composite image, the pixel values were in floating point data format. This prevented further analysis, as Arc Info 9.2 suite supports editing of rasters with only integer DN values. As a result, the pixels of annual composite image had to be converted to integer values. Data accuracy to two places of decimals were lost in the process.

iii) Image Enhancement:

This method of image processing improved the quality of an image to facilitate better information extraction and image interpretation. Two types of image enhancements are applied on satellite images: the point operations, also referred to as radiometric corrections, which change value of individual pixels irrespective of other surrounding pixels and local operations, also known as spatial enhancements, which changes the value of a pixel in the context of values from neighbouring pixels (Bhatta, 2008).

In this research the contrast of the DMSP-OLS images was enhanced to facilitate visual interpretation. The method of contrast enhancement was used for the purpose. It is one of the most commonly used methods of image enhancements. Standard deviation stretch was applied on the images. The method of standard deviation stretch works according to the following formula:

$$y = \frac{S_y}{S_x}(x - x_m) + y_m$$

Where

x= original input data

y= converted output data

X_m= average of input image

S_x= Standard deviation of input image

S_y= Standard deviation of output image

Y_m= average of output image

Equation 4.3: Standard Deviation stretch

After pre - processing, the images were analyzed to obtain mean and standard deviation of light information for each spatial scale. The steps of image analyses are described in the next section.

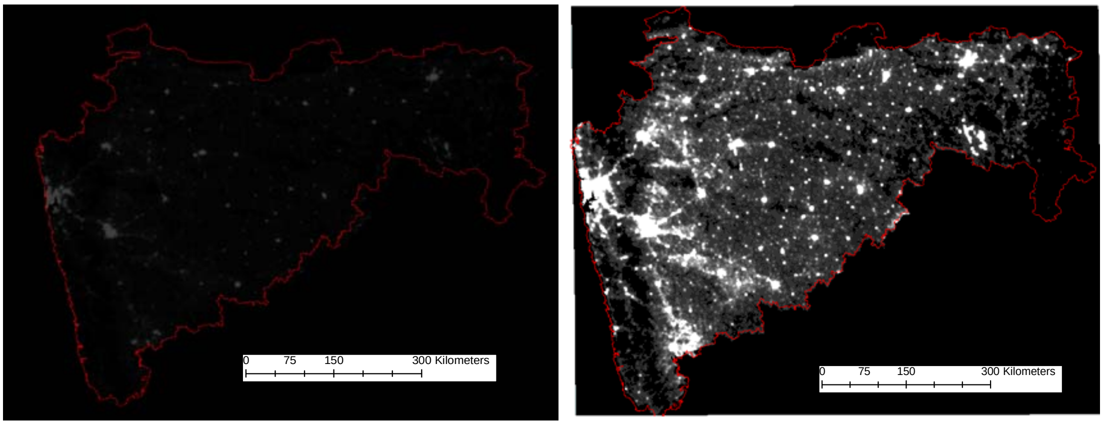


Figure 4.8: Effect of contrast enhancement on the DMSP-OLS image: (a) No stretch; (b) 2 Standard Deviation Stretch

4.6.4: Image Analysis:

The different image analyses undertaken in this research are:

i. Calibration of radiance for fixed gain images;

ii. Calculation of mean and standard deviation of radiance, brightness and stable lights.

i) Calibration of radiance of fixed gain images:

Radiance is defined as the amount of radiant energy, or upward-welling energy, per area per angle per wavelength interval. It is recorded in $W/cm^2/sr/\mu m$. The radiance values were calibrated using radiance per DN information obtained from NGDC/NOAA. These calibration values were specific for each gain setting. The table with radiance per DN values for all the gain settings are given in Appendix (table 10.5). The gain 20 image was multiplied with the radiance per DN value for gain 20 (7.62E-09 $W/cm^2/sr/\mu m$) and the gain 50 image was multiplied by radiance per DN value for gain 50 (2.41E-10 $W/cm^2/sr/\mu m$). As the resultant values were very small, for the convenience of calculation, the radiance values were multiplied by a constant factor of 10^9. The saturation radiances were 4.80E-07 $W/cm^2/sr/\mu m$ for gain 20 dB images and 1.52E-08 $W/cm^2/sr/\mu m$ for gain 50 dB images. The pixels with DN values of 63 and above were saturated. Methods on the conversion of pixel values to radiance values were discussed and confirmed by NGDC/NOAA by email communication (Tuttle, 2009).

Each pixel in the annual composite image of 2001 represented a brightness value. As a result, no further calibration was applied to that image. According to the information obtained from personal communication with NGDC (Tuttle, 2009), this image had not been further processed to radiance, hence the values obtained from it were reported as absolute brightness ($Watts/cm^2/\mu m$) and not as radiance.

ii) Calculation of mean and standard deviation of radiance, brightness and stable lights for districts, taluks and villages:

Mean and standard deviation of radiance, brightness and stable lights were calculated over the spatial units using the following equations:

$$Average = \frac{Total\,Radiance/\,brightness/\,stable\,lights}{Area\,(Km^2)}$$

Equation 4. 4: Mean of Radiance/Brightness/Stable lights

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Where,

σ = standard deviation

N = number of spatial units

x_i = radiance, brightness and stable lights over the spatial units

$\bar{x}$ = average

Equation 4. 5: Standard deviation of radiance/brightness/stable lights

The average radiance and brightness obtained for each areal units were correlated with the shortlisted demographic and amenities data obtained from Primary Census Abstract of Maharashtra. Out of all the census variables, 38 were selected through different statistical tests (section 4.2.3). They were correlated with mean and standard deviation obtained from image analyses (section 4.3.5). The correlation coefficients were calculated by bootstrapping. The final steps in the selection of census metrics are explained in section 4.5. The next section describes the different data quality issues that were encountered while dealing with the satellite images and the census datasets.

4.7: Data quality issues

This section describes the data quality issues encountered while combining the three types of data used in this study. Spatial data quality issues are important considerations in the process of production and distribution of maps (Chapman, 2004, Goodchild, 2009, Hunter and Goodchild, 1995, Comber et al., 2006). The Commission of Spatial Data Quality was set up in 1991 by the International Cartographic Association to address these issues (Guptill and Morrison, 1995). Broadly, spatial data quality issues relate to aspects of data accuracy, spatial and temporal resolution, data completeness and data consistency (Lillesand et al., 2004, Guptill and Morrison, 1995, Hunter et al., 2003a). Spatial data quality generally consists of: (1) positional accuracy, (2) attribute accuracy, (3) logical consistency, (4) data completeness and (5) data lineage (Morrison 1995, Hunter et al. 2003a, Comber et al. 2006, Reinke and Jones 2006). These key spatial data quality issues have already been summarized in many international guidelines such as those presented by the United Nations Environmental Programme, the International Organization of Standardization, the United States National Committee on Digital Cartographic Standards and Spatial Information Council of Australia and New Zealand guidelines (formerly the Australia New Zealand Land Information Council).

The data quality issues encountered in this research specifically relate to positional accuracy, attribute accuracy, absence of metadata and data lineage. The details of these elements are explained in the following sections:

4.7.1: Positional accuracy:

Positional accuracy refers to the spatial or geographic precision of a dataset. There are two types of positional errors. The first type of error may be calculated with respect to known location of points and is known as absolute accuracy error. The second type of positional accuracy is called the relative accuracy and is considered with respect to another dataset (McCoy, 2005, Reinke and Jones, 2006). Positional accuracy errors can occur during data acquisition and may also occur during data processing and transformation.

Positions of spatial datasets are also defined in terms of projections and coordinate systems. Different projections and coordinate systems introduce different proportions of bias and distortion in the spatial properties of a dataset (Drummond, 1995). The coordinates of points differ on the basis of datums used to record them. A datum is a reference frame used to define the coordinates of points on the earth surface. In India, a new map policy was introduced in 2005 (Survey of India, 2005). In the new policy, two map datums were proposed: one for defence series maps and the other for civilian use called open series maps. The open series maps were produced in Universal Transverse Mercator (UTM) projection system on WGS 1984 datum (Ghosh and Dubey, 2008). The WGS 1984 datum is a geocentric ellipsoid (defined with the centre of the earth as its origin) put forward by the US DoD and is a conventional terrestrial system (Ghosh and Dubey, 2008).

When datasets are spatially located, there occurs some degree of positional inaccuracy, the significance of which is determined by the study, otherwise referred to as fitness-of-use. This effect is further compounded when multiple data sources are used (Thomlinson et al., 1999, Elmore et al., 2000, Armston et al., 2002, Coulter and Stow, 2009) and it becomes necessary to know the relative positional accuracy between different datasets to draw valid comparisons (Means et al., 1999, Armston et al., 2002, McCloy, 2006, Reinke and Jones, 2006, Weber et al., 2008). Points of known spatial position or Ground Control Points (GCPs) are used to assess the absolute positional accuracy of datasets (Reinke and Jones, 2006, Thomlinson et al., 1999). The accuracy is commonly measured in terms of Root Mean Square Error (RMSE).

In this research, two types of positional errors were noted. The first was error in relative positional accuracy between DMSP-OLS images and the SAC datasets. For this, the positional inaccuracy between satellite images and the spatial datasets was rectified by the method of geometric correction. This is explained in detail in satellite image processing (section 4.3.4.2). The second type of positional error was the inconsistency between district and village datasets obtained from SAC. SAC provided datasets defining the boundaries of districts, taluks and villages of the state of Maharashtra. A district is divided into a number of taluks and the latter into villages. Therefore, the boundaries of the bordering taluks are coincident with the district boundaries and those of bordering villages are coincident with the taluk boundaries. However in the dataset, the boundaries of bordering villages were not overlying on the taluk boundaries. There was a gap of 300 – 500 m between these two boundaries across the dataset. Fig 4.9 illustrates the positional inaccuracy between the taluk and the village datasets. The red line shows the taluk boundaries while the black lines demarcate the villages.

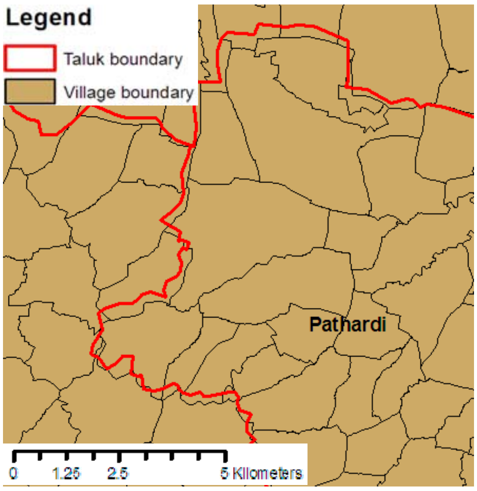


Figure 4.9: Positional inconsistency between taluk and village datasets

The probable reason for this error might be the difference in the source datasets that were used to prepare them. However, as the regions at different spatial scales were treated separately in this research, this error did not have significant impact on the results. Also, the deviation in the boundaries of the datasets was less than 500 m (half the pixel size of the satellite image). As a result it was assumed to have little or no effect on the mean pixel values obtained from the satellite images which were used to build the models.

4.7.2: Attribute Accuracy

Attribute accuracy can have a significant impact on the quality and usability of a dataset. Attributes make up the thematic component of a dataset. They refer to the 'facts' about different features in a dataset and helps to describe them. Goodchild (1995) suggested that the range of values of an attribute uncertainty results from repeated measurements of the same feature. There are two terms commonly associated with attribute uncertainty: precision and accuracy. Precision refers to the level of reported detail for an attribute while accuracy refers to the difference between the spatial data and ground data (Drummond, 1995, Goodchild, 1995, McCloy, 2006). Positional accuracy sometimes influences attribute accuracy via the positional matching between spatial features and their attributes (Goodchild, 1995, Reinke and Jones, 2006).

In this research, a number of errors associated with attributes of the taluks and village datasets were noted. These errors can be broadly divided into two types:

I. Inter - dataset error;
II. Intra - dataset errors.

I) Inter – dataset error was noted between census data and SAC datasets.

There were errors in the numbers of taluks and villages between the census and the SAC datasets. Census data for the state of Maharashtra accounted for 353 taluks and 41,095 villages while the SAC dataset had defined 317 taluks and 43,669 villages (a difference of 10%). Table 4.5 shows percentage of difference in the number of taluks in the two datasets. Some districts such as Amravati, Wardha, Thane, Solapur, Satara, Sindhudurg and Kolhapur had the same number of taluks recorded in both the datasets. There were 36 missing taluks in the SAC dataset with maximum absolute number of taluks missing for the district of Latur (five taluks missing i.e. 50% of the total number of taluks), Gadchiroli (four missing taluks) and Bid (three missing taluks). The remaining districts had one taluk missing in each of them. In terms of percentage, Parbhani had 44% of the taluks missing and Jalna had 37.5% of the taluks missing in the SAC dataset.

There was a huge difference in the number of villages for each district between the census and the SAC database. The least difference was noted for the district of Nagpur where the difference in the number of villages between the two datasets was 41. Districts such as Dhule, Washim, Bhandara, Nashik, Bid and Osmanabad have a difference of more than 1000 villages between the two datasets. More than 100% difference in the number of villages is found in the districts of Kolhapur, Latur, Osmanabad and Washim The difference in percentage between the two datasets in the number of villages for each district is shown in table 4.5.

Table 4. 5: Difference in the number of taluks between the census and the SAC dataset

DISTRICTS	**Number of taluks According to census**	**Number of taluks According to SAC**	**Difference (%)**
Ahmadnagar	14	13	7.14
Akola	7	7	0.00
Amravati	14	14	0.00
Aurangabad	9	10	11.11
Bhandara	7	6	14.29
Bid	11	8	27.27
Buldana	13	12	7.69
Chandrapur	14	10	28.57
Dhule	4	5	25.00
Gadchiroli	12	8	33.33
Gondiya	8	7	12.50
Hingoli	5	3	40.00
Jalgaon	15	14	6.67
Jalna	8	5	37.50
Kolhapur	12	12	0.00
Latur	10	5	50.00
Mumbai	0	1	100.00
Mumbai (Suburban)	0	3	100.00
Nagpur	14	13	7.14
Nanded	16	8	50.00
Nandurbar	6	7	16.67
Nashik	15	13	13.33
Osmanabad	8	6	25.00
Parbhani	9	5	44.44
Pune	14	17	21.43
Raigarh	15	14	6.67
Ratnagiri	9	10	11.11
Sangli	9	8	11.11
Satara	11	11	0.00
Sindhudurg	8	8	0.00
Solapur	11	11	0.00
Thane	15	15	0.00
Wardha	8	8	0.00
Washim	6	6	0.00
Yavatmal	16	14	12.50
Total	**353**	**317**	10.20

Table 4.6: Difference in the number of villages between the census and the SAC dataset

DISTRICTS	Number of Villages According to census	Number of Villages According to SAC	Difference (%)
Ahmadnagar	1578	1379	12.61
Akola	862	736	14.62
Amravati	1679	1612	3.99
Aurangabad	1300	1580	21.54
Bhandara	778	1902	144.47
Bid	1354	185	86.34
Buldana	1297	1527	17.73
Chandrapur	1472	1163	20.99
Dhule	678	1728	154.87
Gadchiroli	1521	732	51.87
Gondiya	893	1586	77.60
Hingoli	672	967	43.90
Jalgaon	1491	1950	30.78
Jalna	963	1280	32.92
Kolhapur	1196	0	100.00
Latur	921	1861	102.06
Mumbai	0	1791	100.00
Mumbai (Suburban)	0	1439	100.00
Nagpur	1628	1587	2.52
Nanded	1546	1348	12.81
Nandurbar	935	0	100.00
Nashik	1923	722	62.45
Osmanabad	729	1866	155.97
Parbhani	830	1514	82.41
Pune	1844	2118	14.86
Raigarh	1859	2035	9.47
Ratnagiri	1539	0	100.00
Sangli	721	0	100.00
Satara	1716	0	100.00
Sindhudurg	743	0	100.00
Solapur	1138	1672	46.92
Thane	1727	936	45.80
Wardha	1004	1536	52.99
Washim	702	1837	161.68
Yavatmal	1856	1215	34.54
Subtotal	**41095**	**41804**	0.00
Villages with no district codes	0	1802	-
Creeks attributed as villages	0	63	-
Total	**41095**	**43669**	6.26

II) Intra-dataset error was noted within the village database. Four types of intra – dataset errors were noted:

1. Missing district codes[8] for the villages:

   1802 villages in the SAC dataset have no district codes attached to them. In other words, from the village dataset it was not possible to identify the districts they belong to. This error was noted for the villages of the districts of Nandurbar, Satara, Ratnagiri, Sangli, Sindhudurg and Kolhapur. The error was rectified after manual comparison of the SAC and the census datasets.

2. Error in district codes of the villages:

   a. Upon comparison with the census database, it was further noted that none of the villages were correctly attributed with their respective district codes. Table 4.7 shows the actual location of the villages according to the census and their location denoted in the SAC database.

   b. Mumbai and Mumbai (suburban) districts are urban in nature. They form part of the Mumbai Metropolitan Area (MMA) and do not have any rural settlement or villages in them. The SAC database, however, accounts 1791 and 1439 villages respectively in each of them.

3. Error in thematic representation: Sixty three creeks in the district of Sindhudurg were denoted as villages in the SAC dataset.

Table 4.7: Difference in the location of villages between census and SAC dataset

Original location of villages in districts (census)	Districts where same Villages are demarcated in SAC dataset
Ahmadnagar	Wardha
Akola	Sindhudurg
Amravati	Nandurbar/Dhule
Aurangabad	Nanded
Bhandara	Pune
Bid	Parts of Nagpur
Buldana	Ratnagiri
Chandrapur	Solapur
Dhule	Thane
Gadchiroli	Sangli
Gondiya	Satara
Hingoli	Jalna
Jalgaon	Raigarh
Jalna	Bid
Kolhapur	NA

[8] District code is a code generated by census unique for each district. E.g. district Nandurbar has district code 01

Latur	NA
Mumbai	Akola and Washim
Mumbai (Suburban)	Buldana and parts of Jalgaon
Nagpur	Ahmadnagar
Nanded	Aurangabad
Nandurbar	NA
Nashik	Osmanabad
Osmanabad	Chandrapur
Parbhani	Parts of Parbhani and Hingoli
Pune	Yavatmal
Raigarh	Amravati
Ratnagiri	NA
Sangli	NA
Satara	NA
Sindhudurg	NA
Solapur	Gadchiroli
Thane	Latur and some parts of Bid
Wardha	Jalgaon
Washim	Nashik
Yavatmal	Kolhapur

Because of the presence of the intra dataset errors, it was labourious and time-consuming to combine the results obtained for the villages from the DMSP-OLS images with the SAC datasets. After identifying these errors, the village database was corrected and all the villages properly attributed with their respective district codes as obtained from the census.

4.7.3: Other data quality issues:

Apart from these major data accuracy issues, some other data quality issues were also noted in the SAC dataset. They include an absence of metadata and lineage documentation.

Data lineage refers to the description of the acquisition, compilation and any other transformations or corrections applied to the dataset (Clarke and Clark, 1995). For example, information on data source, methods of data collection, map projections, datum, data correction and calibration are included in data lineage (Clarke and Clark, 1995, Morrison, 1995, Reinke and Jones, 2006). Although information on map projections and datum were present in the SAC dataset, no further information on data source and calibration were present.

Metadata is "data about data". It includes descriptions and measurements of all the data quality elements of any dataset such as the methodology used to collect the data, limitations present in the dataset and information on spatial and attribute accuracy (Hunter et al., 2003b, Comber et al., 2006). There was no metadata attached to the dataset and hence there was absence of any prior information regarding the usability of the datasets (Comber et al., 2006).

4.8: Combining results from census and satellite data processing

4.8.1: Correlation and process of bootstrapping:

Correlation between the 38 selected census metrics and mean of brightness and stable lights over the districts were calculated. As there were only 24 districts, the correlation coefficients were calculated using the method of bootstrapping to overcome any error occurring due to small sample size.

The method of bootstrapping was first proposed by Bradley Efron (Chernick 2008b; Stanford University 2004) for independent and identically distributed observations. Bootstrapping is a resampling procedure (Chernick 2008b; Hesterberg et al. 2005) practised in statistics. Other resampling methods include jacknife and permutation tests (Chernick 2008c). Efron combined the methods of jacknife and delta to come up with the concept of bootstrapping (Chernick 2008c). The shape of bootstrap distribution is nearly normal and the mean is close to the mean of the original sample. Unlike other methods of resampling, process of bootstrapping can be applied to any sample size (Chernick 2008a) making it particularly suitable for this research.

The bootstrapping of correlation coefficients was carried out sequentially by the following steps:

1. A random sample was drawn from the Population (i.e. all the 32 districts). This sample consisted of 80% of the Population i.e. 24 districts were chosen in the first step for the purpose of reserving data for validation.
2. The bootstrap sample used a simple random sample with replacement selected from the main sample (obtained from step 1). 1000 bootstrap samples were drawn from the original sample for this study. Although 50 – 200 bootstrap samples are considered sufficient for large Population (Chernick, 2008b, Chernick, 2008a), in this case 1000 bootstrap samples were drawn to overcome any impact of small sample size of 24 districts.
3. The correlation between mean and standard deviation of brightness and stable lights and the chosen 38 census metrics of the 24 districts were calculated by bootstrapping from the 1000 resamples at the 95% confidence interval.

4. Bias and standard error of all the correlations was noted. Variables with a bias[9] of less than 0.05 and standard error[10] of less than 0.2 were chosen.

The results from bootstrapping of all the coefficients are given in table 10.8 in the appendix. Ten variables were finally chosen out of the 38 variables for proposing models in this research. The selected variables are:

- *Number of households per square kilometre*
- *Total population density*
- *Urban population density*
- *Female literates per square kilometre*
- *Total number of workers per square kilometre*
- *Percentage of households with car, jeep and van*
- *Percentage of households with access to electricity as power source*
- *Percentage of households with television*
- *Percentage of permanent houses*
- *Per Capita District Domestic Product*

The results of bootstrapping of correlations of the selected metrics are presented in table 4.8. From table 4.8, it was noted that the bias varies from -0.05 to 0.02 for all the census metrics. Negative bias was noted for all the variables when correlated with mean and standard deviation of stable lights and brightness. The standard errors were less than 0.2 for most of the variables.

Table 4.8: Results from bootstrapping correlation coefficients of 10 selected variables

		Mean stable lights	**Mean brightness**	**Standard deviation stable lights**	**Standard deviation brightness**
Number of	Pearson Correlation	.798	.792	.573	.638

[9] Bootstrap estimate of bias refers to the difference between the mean of the bootstrap distribution and the value of the statistic i.e. correlation coefficient in this case, in the original sample.

[10] The bootstrap standard error is the standard deviation of its distribution in the bootstrap sample

	Sig. (2-tailed)			.000	.000	.003	.001
	Bootstrap	Bias		-.021	-.020	-.021	-.014
		Std. Error		.123	.129	.159	.173
		BCa the 95% confidence Interval	Lower	.550	.520	.207	.197
			Upper	.910	.921	.772	.875
Total population density	Pearson Correlation			.830	.800	.623	.638
	Sig. (2-tailed)			.000	.000	.001	.001
	Bootstrap	Bias		-.017	-.009	-.018	-.003
		Std. Error		.098	.094	.142	.143
		BCa the 95% confidence Interval	Lower	.621	.586	.266	.264
			Upper	.923	.916	.807	.876
Female literates per square kilometre	Pearson Correlation			.767	.798	.556	.669
	Sig. (2-tailed)			.000	.000	.005	.000
	Bootstrap	Bias		-.026	-.029	-.027	-.028
		Std. Error		.131	.131	.159	.170
		BCa the 95% confidence Interval	Lower	.439	.392	.209	.215
			Upper	.901	.916	.756	.864
Total number of workers per square kilometre.	Pearson Correlation			.761	.681	.532	.492
	Sig. (2-tailed)			.000	.000	.007	.015
	Bootstrap	Bias		-.014	.004	-.019	.019
		Std. Error		.118	.120	.167	.165
		BCa the 95% confidence Interval	Lower	.482	.401	.156	.164
			Upper	.892	.874	.748	.802
Percentage of households with car, jeep and van	Pearson Correlation			.687	.807	.414	.650
	Sig. (2-tailed)			.000	.000	.045	.001
	Bootstrap	Bias		-.007	-.022	.004	.001
		Std. Error		.139	.122	.153	.155
		BCa the 95% confidence Interval	Lower	.085	.405	-.189	.148
			Upper	.901	.936	.743	.928
Percentage of	Pearson Correlation			.789	.882	.622	.785

	Sig. (2-tailed)			.000	.000	.001	.000
	Bootstrap	Bias		-.011	-.016	-.003	-.008
		Std. Error		.092	.069	.107	.095
		BCa the 95% confidence Interval	Lower	.545	.725	.330	.493
			Upper	.906	.945	.794	.909
Percentage of permanent houses	Pearson Correlation			.633	.654	.406	.474
	Sig. (2-tailed)			.001	.001	.049	.019
	Bootstrap	Bias		-.004	-.007	.001	.013
		Std. Error		.111	.124	.122	.153
		BCa the 95% confidence Interval	Lower	.340	.370	.058	.115
			Upper	.835	.862	.648	.831
Percentage of households with access to electricity as power source	Pearson Correlation			.454	.483	.339	.402
	Sig. (2-tailed)			.026	.017	.105	.051
	Bootstrap	Bias		-.008	-.002	-.004	-.001
		Std. Error		.251	.161	.258	.134
		BCa the 95% confidence Interval	Lower	-.147	.117	-.220	.120
			Upper	.832	.753	.768	.655
Urban population density	Pearson Correlation			.819	.963	.645	.890
	Sig. (2-tailed)			.000	.000	.001	.000
	Bootstrap	Bias		.012	-.007	.018	-.003
		Std. Error		.050	.034	.065	.081
		BCa the 95% confidence Interval	Lower	.658	.854	.459	.573
			Upper	.945	.991	.829	.987
Per Capita District Domestic Product (INR 1998 – 99)	Pearson Correlation			.543	.812	.322	.812
	Sig. (2-tailed)			.006	.000	.126	.000
	Bootstrap	Bias		-.020	-.052	-.011	-.047
		Std. Error		.254	.201	.232	.177
		BCa the 95% confidence Interval	Lower	-.164	.232	-.319	.301
			Upper	.900	.943	.708	.923

4.9: Summary:

This chapter describes the methods undertaken to prepare the census data and the satellite images for further analyses. 24 districts were randomly sampled from the state of Maharashtra. Statistical tests were conducted on census datasets to test the normality of distribution. Satellite images obtained from

DMSP-OLS were analyzed to calculate mean and standard deviation of stable lights, radiance and brightness over regions of varying spatial scales. The census metrics which were distributed normally over the sampled districts were correlated with the results from satellite image processing by the using the method of bootstrapping. Finally a list of 10 census metrics was shortlisted to propose models. Some of these metrics such as number of female literates per square kilometre and percentage of permanent census houses were normally distributed over the study area. In addition to being statistically fit for selection, these metrics were also important parameters for determining the overall development of an area. Other metrics such as total population per square kilometre and PCDDP were important components of any census. All these issues were taken into consideration for the final selection of the ten metrics. The details of the models at different spatial scales using these metrics are explained in the later chapters.

Key findings from the chapter:

- The details of the method followed in the research are described in this chapter.
- Major statistical tests included sampling, tests of normal distribution and bootstrapping the correlations.
- Mean and standard deviation of stable lights and brightness were obtained from satellite image processing.
- 10 census metrics (both demographic and socio-economic) were shortlisted to propose surrogate census.

5. Scope and Limitations of creating a surrogate census from fixed gain radiance calibrated images

[This chapter has been published as the following peer reviewed publication:
Roychowdhury, K, Jones, SD, Arrowsmith, C & Reinke, K 2011, 'A comparison of high and low gain DMSP/OLS satellite images for the study of socio-economic metrics', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 4, no. 1, pp. 35-42.]

This chapter describes the usefulness of fixed gain radiance calibrated DMSP-OLS night-time images in proposing a surrogate census. Two fixed gain images were used: a low gain image with gain setting of 20 dB and a high gain image with gain setting of 50 dB. Firstly, the chapter introduces the method undertaken to obtain the results (section 5.2) followed by analyses of correlations between radiance and census metrics (section 5.3). In section 5.4 the models are described. The proposed models are further examined in detail and validated over the withheld districts in section 5.5. The final models are proposed in section 5.6. The chapter concludes with a section on the selection of optimum gain (section 5.7) and the limitations of single orbit fixed gain images (section 5.8). This latter section explains the reasons for not using these images further in the research.

5.1 Introduction:

Fixed gain images are captured by the OLS sensor only on some selected dates of a year. Therefore the availability of these datasets is limited both in terms of their numbers and temporal frequency. For the state of Maharashtra, two fixed gain images, one with gain setting 20dB and the other with gain 50 dB were chosen from DMSP satellite F15 for analyses. The images were all 8 bit (pixel values from 0 – 255). The gain 20 db image was captured on 17th September, 2001 and the gain 50 dB image was collected on 14th September of the same year.

5.2 Method:

The chosen images (the G20 and the G50) were calibrated to radiance using the information on the radiance per digital number (DN) values for individual gain settings provided by the National Geophysical Data Centre (NGDC) with the images. The DMSP-OLS images were reprojected to the Lambert Conformal Conic Projection (datum WGS84), common to the vector datasets obtained from India. Images were converted to grid to facilitate comparisons with the socio economic dataset. However, the original image resolution (approximately 833m) was maintained during the conversion in order to prevent any loss of information. The images were geo-referenced with respect to the vector datasets. Average and standard deviation of radiance obtained for the 35 districts of Maharashtra were correlated with the demographic and amenities datasets from the last Indian census of 2001. The results were then compared for images with both the gain settings to decide on the appropriate gain. Ten shortlisted metrics as obtained from the statistical tests (explained in section 4.8) were used for the analyses.

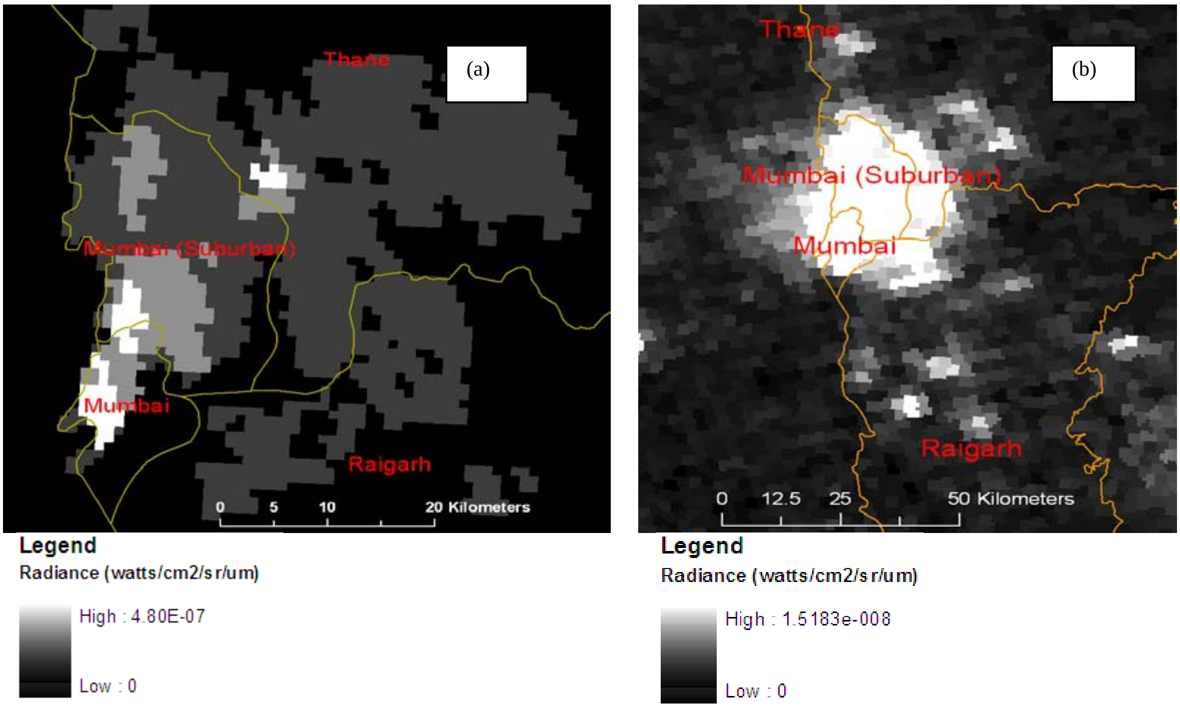


Figure 5.1: Gain 20 (a) and Gain 50 (b) images showing parts of Maharashtra. Gain 20 image shows variation of radiance within urban areas (Mumbai in this figure) and Gain 50 image captures more light. This image tends to exaggerate urban boundaries, an effect known as blooming, which can be easily noticed on the coast along Mumbai and Mumbai suburban regions in the figure.

The analyses started with examining correlations among the chosen census metrics and mean and standard deviation of radiance obtained from gain 20 dB and gain 50 dB images. To achieve this, correlation matrices were calculated for each district to identify the degree of correlation between census metrics and recorded radiance information from night-time images.

5.3.1 Correlations with gain 20 dB image:

All the census metrics exhibited positive correlations with both mean and standard deviation of radiance obtained from gain 20dB image. The correlations with mean radiance and census metrics are shown in figure 5.2 and those with standard deviation radiance are illustrated in figure 5.3.

Weak correlations were found for most of the census metrics. The correlation coefficients (r values) with mean radiance ranged from 0.08 for *percentage of households with access to electricity* to 0.53 for urban population density. There were no significant correlations at the 95% confidence interval for *number of female literates per square kilometre, total workers per square kilometre, percentage of households with access to electricity* and *PCDDP* (marked with blue boxes in figure 5.2). Pune and Aurangabad were noted as outliers in the correlations with all the census metrics. These two districts had high mean radiance from gain 20 dB images (more than 10 watts/cm^2/μm/sr). Some of the districts such as Bhandara, Ratnagiri, Sindhudurg, Solapur, Wardha and Washim exhibited no mean radiance (0 watts/cm^2/μm/sr) from gain 20 dB image. As a result all these districts were distributed in a linear cluster in the graphs. These are circled with green in the figure 5.2. Some other smaller clusters of districts were also noted from the correlations. Since the mean radiance obtained from gain 20 dB image ranged from 0 to 13 watts/cm^2/μm/sr over the 24 sampled districts, not much variation was noted in the distribution over the region.

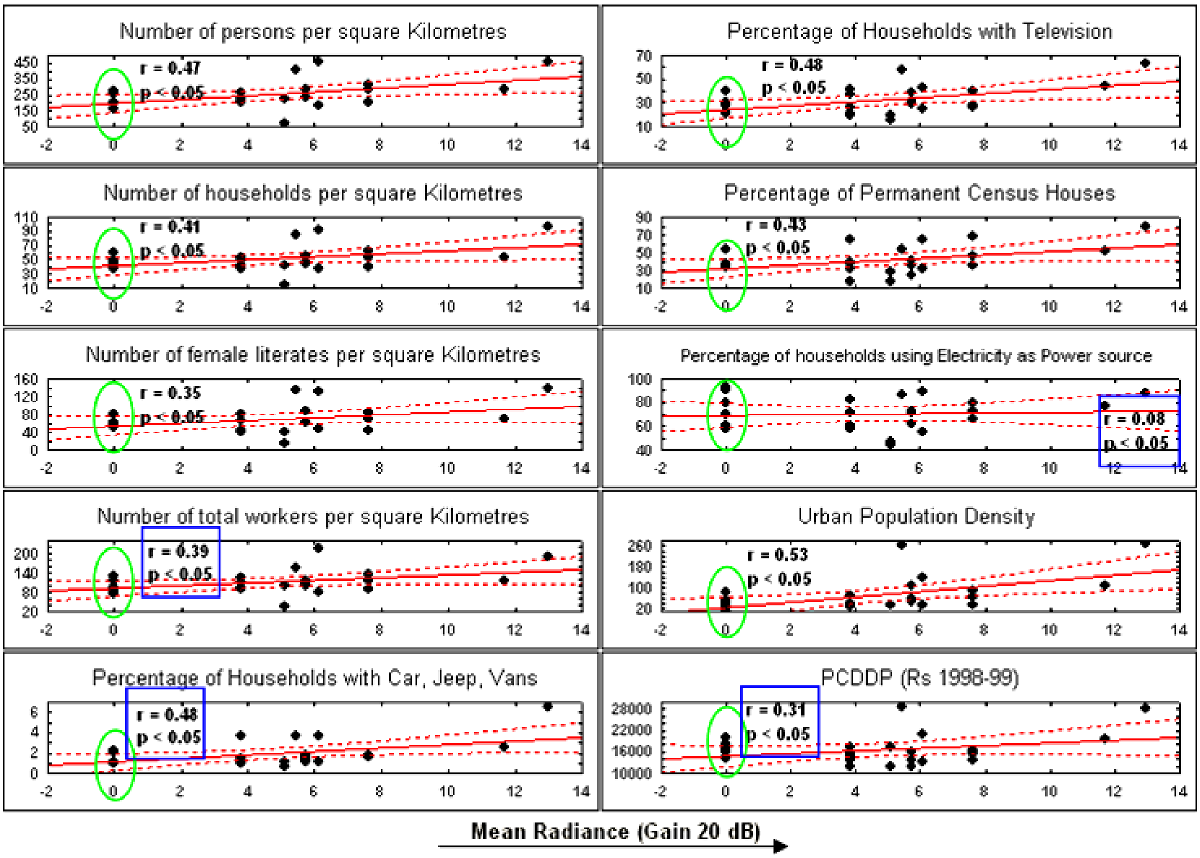


Figure 5.2: Correlations between census metrics and mean radiance as recorded from gain 20 dB image

From the correlation matrices calculated with standard deviation of radiance obtained from gain 20 dB image and the census metrics it was found that correlation coefficients were not significant for most of the metrics at the 95% confidence interval. *Number of households per square kilometre, total population per square kilometre, female literates per square kilometre, total workers per square* *kilometre, percentage of permanent census houses, percentage of households using electricity as power source* and *PCDDP* did not have significant correlations with standard deviation radiance from gain 20 dB image. All the metrics which did not exhibit significant correlations are marked with blue boxes in figure 5.3. Pune was noted as an outlier district in all the correlations. The standard deviation of radiance for Pune was more than 12 watts/cm^2/μm/sr. Some of the districts such as Bhandara, Ratnagiri, Sindhudurg, Solapur, Wardha and Washim have zero standard deviation of radiance and therefore appear as a linear cluster in the graphs (demarcated with green circles in figure 5.3).

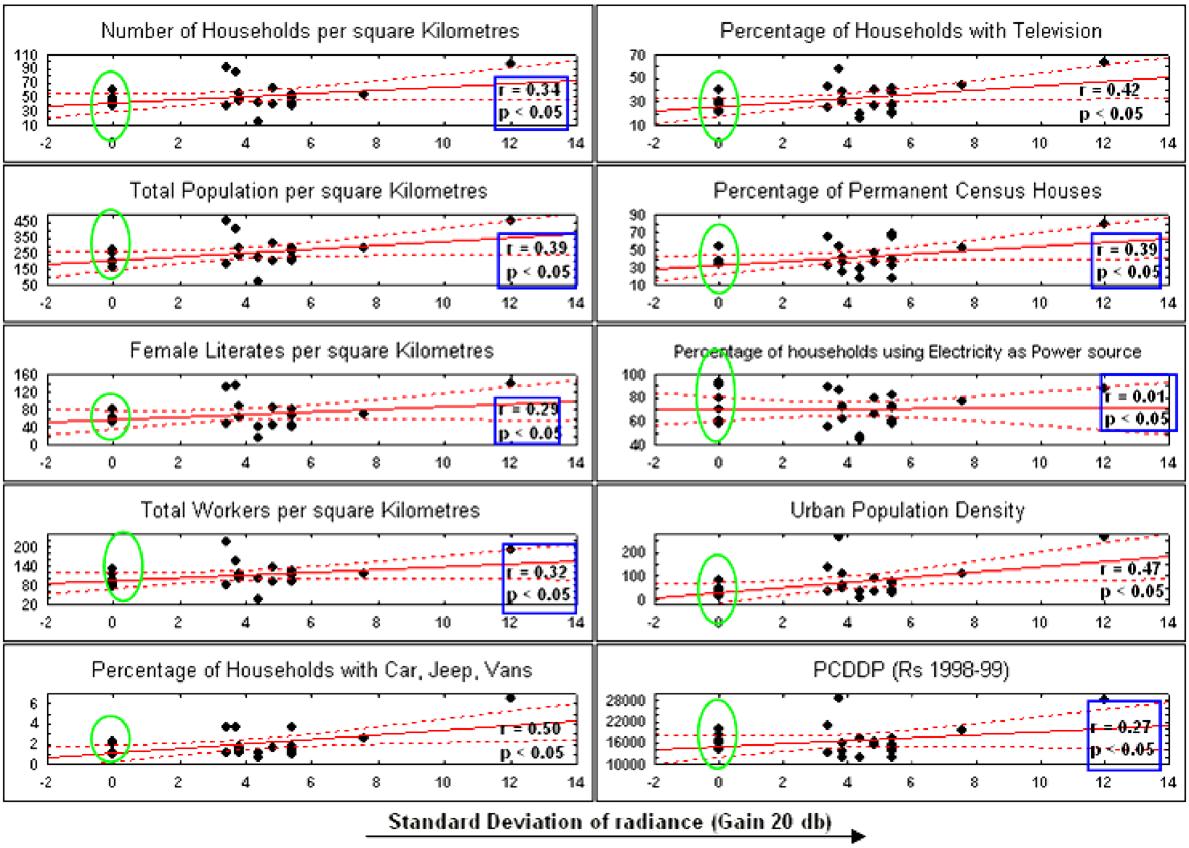


Figure 5.3: Correlations between census metrics and standard deviation of radiance as obtained from gain 20 dB image

Overall, the census metrics exhibit weak correlations with mean and standard deviation of radiance as obtained from gain 20 dB image. The characteristic of low gain setting DMSP-OLS images is one of the probable reasons for weak correlations. These images capture the variation in radiance over small areas but fail to detect lights from smaller sources such as rural settlements in the districts. As a result mean and standard deviation of radiance from gain 20 dB image were recorded only for districts with major urban centres such as Pune, Aurangabad and Nagpur.

In contrast to the gain 20 dB images, census metrics showed stronger correlations with mean and standard deviation of radiance obtained from gain 50 dB image. The results of correlations with gain 50 dB image are explained in the following section.

5.3.2 Correlations with gain 50 dB image:

There were positive correlations significant at the 95% confidence interval between the chosen census metrics and mean and standard deviation radiance as obtained from gain 50 dB image. The correlations with mean radiance are shown in figure 5.4 and with standard deviation radiance are illustrated in figure 5.5.

Moderate correlations were noted between the census metrics and mean radiance. The correlation coefficients calculated at the 95% confidence interval ranged from 0.55 for *PCDDP* to 0.82 for each of *total population per square kilometre* and *percentage of households with cars, jeeps and vans*. All the correlations were significant at the 95% confidence interval. Although the districts of Pune and Kolhapur were noted as outliers in some of the correlations, a general trend in the distributions was noted for all the metrics. Therefore, these two districts did not seem to have a strong impact on the values of correlation coefficients.

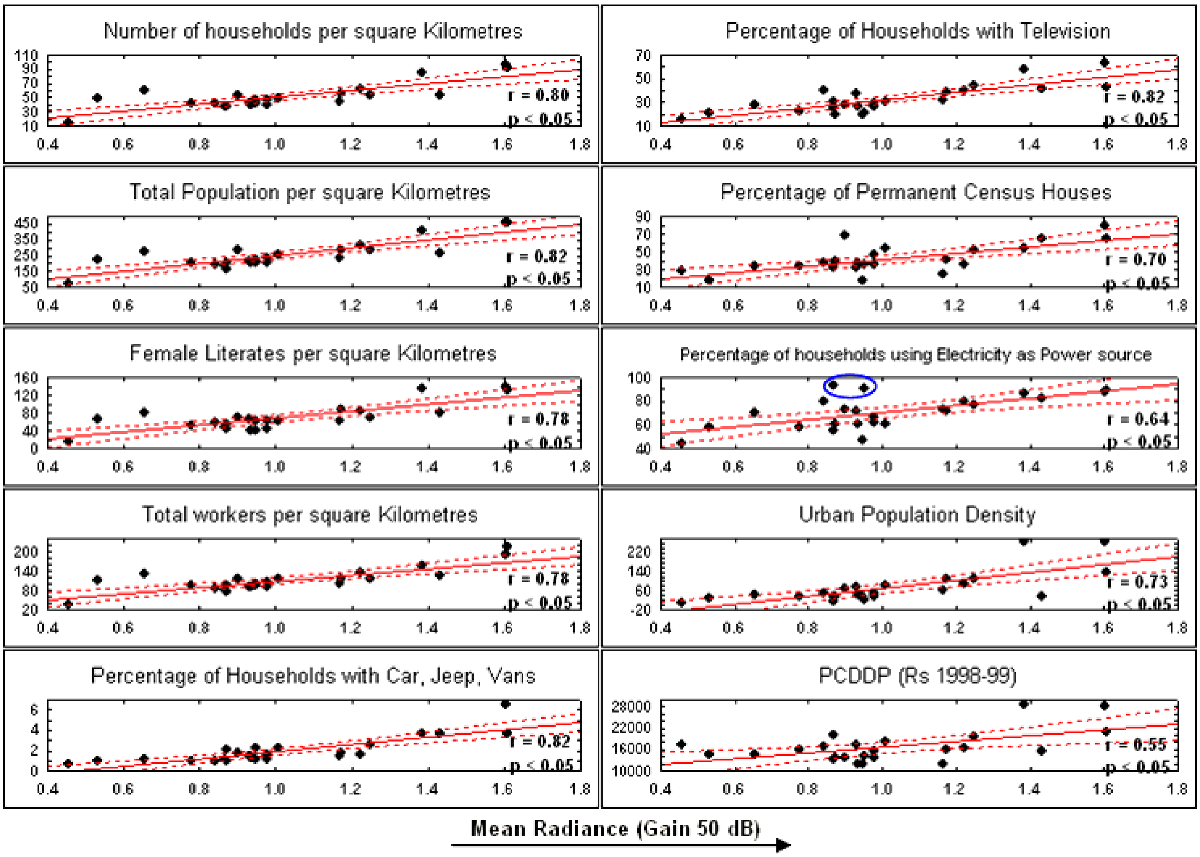


Figure 5.4: Correlations between census metrics and mean radiance as obtained from gain 50 dB image

Some of the districts were scattered when *percentage of households with access to electricity* was correlated with mean radiance. Districts such as Ratnagiri and Sindhudurg (encircled by blue in figure 5.4) have more than 90% of the households using electricity as power source. Districts such as Latur, Wardha and Amravati have more than 80% of their households with electricity. However these districts fail to record as high a mean radiance as Pune and Kolhapur. One of the probable reasons for low mean radiance in these districts is the absence of any major urban centres. Other reasons might be the higher use of indoor lights and improved infrastructure in the streets which prevent light from escaping into space, and thereby is not recorded by the DMSP-OLS sensor.

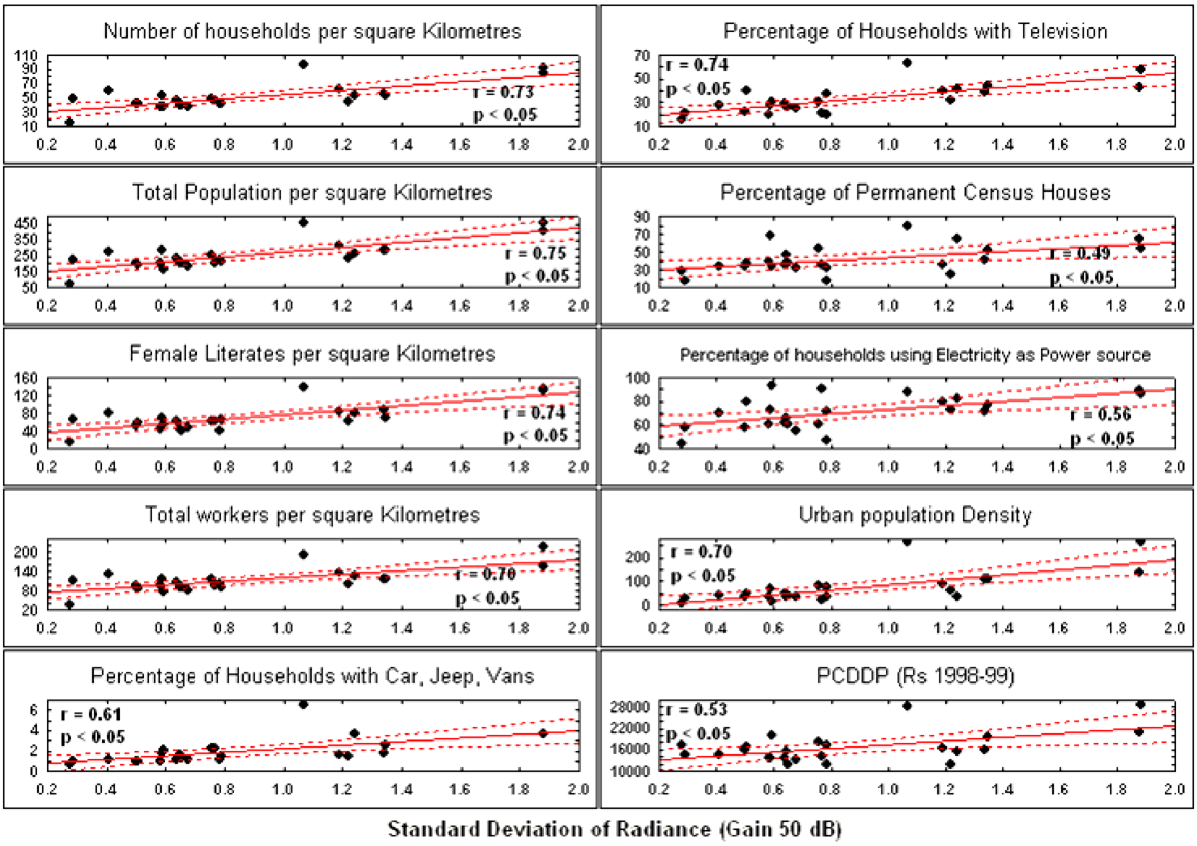


Figure 5.5: Correlations between census metrics and standard deviation radiance as obtained from gain 50 dB image

Correlations with standard deviation of radiance obtained from gain 50 dB image and the chosen census metrics are illustrated in figure 5.5. When compared with correlation coefficients obtained from mean radiance, there were weaker correlations between census metrics and standard deviation of radiance. The correlation coefficients ranged from 0.49 for percentage of permanent census houses to 0.75 for *total population per square kilometre* at the 95% confidence interval. The districts of Kolhapur and Nagpur were noted as outliers in the correlations. The standard deviations of radiance of these districts were more than 1.8 watts/cm^2/µm/sr. These two districts exerted a pulling effect in raising the correlation coefficients for most of the variables. For most of the districts (16 out of 24) the standard deviation radiance ranged from 0.4 to 1.0. These districts are distributed in a cluster in all the graphs showing the correlations. For remaining five districts the standard deviation of radiance ranged from 1.0 to 1.4 watts/cm^2/µm/sr.

Comparing the results obtained from gain 20 dB and gain 50 dB images, it was noted that gain 50 dB image showed more significant correlations with the chosen census metrics. However, to propose the optimum gain setting of DMSP-OLS image for a surrogate census, models were proposed and tested using the radiance obtained from both the gain 20 dB and gain 50 dB images. Two types of models were tested using the fixed gain images: linear regression models with single independent variables or predictors and multiple regression models where two or more independent variables were used. Models with intercept and without intercepts were tested. Models with intercepts have coefficients while models with no intercepts pass through the intersection of the x and y axes. The details of these models and their results are discussed in the following section.

5.4 Development and discussion of models:

5.4.1 Discussion of linear regression models:

Simple linear regression models were calculated using the chosen metrics and mean and standard deviation of radiance as obtained from gain 20 dB and gain 50 dB images. Two types of models were tested: models with intercept and those without intercepts. Adjusted r^2 values of the models are presented in table 5.1. Adjusted r^2 values are considered to be a better coefficient of determination or "goodness of fit" of a given regression model (Zar, 2010). These values are calculated using the following formula:

$$R_a^2 = 1 - \frac{n-1}{n-m-1}(1-R^2)$$

Where, n = number of variables

m = number of independent variables

R^2 = coefficient of determination

Equation 5.1: Calculation of adjusted r^2 values

Table 5.1: Adjusted r^2 values from linear regression models. The highest adjusted r^2 values are shaded in light blue and the lowest adjusted r^2 values are shaded in grey

Census Metrics	**MeanG20 with intercept**	**MeanG20 with no intercept**	**MeanG50 with intercept**	**MeanG50 with no intercept**	**SDG20 with intercept**	**SDG20 with no intercept**	**SDG50 with intercept**	**SDG50 with no intercept**
Number of households per square kilometre	0.13	0.69	0.62	0.96	0.08	0.68	0.50	0.90
Total population per square kilometre	0.18	0.71	0.66	0.96	0.11	0.69	0.55	0.91
Female	0.08	0.66	0.58	0.93	0.04	0.64	0.52	0.90

literates per square kilometre								
Total workers per square kilometre	0.11	0.68	0.60	0.96	0.06	0.67	0.47	0.89
Percentage of households with cars, jeeps and vans	0.19	0.67	0.66	0.85	0.21	0.69	0.34	0.80
Percentage of households with television	0.20	0.72	0.66	0.96	0.14	0.70	0.52	0.90
Percentage of permanent houses	0.15	0.69	0.47	0.94	0.12	0.69	0.20	0.83
Percentage of households using electricity as power source	0.00 (-0.04)	0.62	0.38	0.95	0.00 (-0.05)	0.62	0.29	0.85
Urban population per square kilometre	0.25	0.63	0.52	0.70	0.18	0.59	0.47	0.75
PCDDP (Rs 1998 – 99)	0.05	0.66	0.27	0.94	0.03	0.66	0.25	0.85

The highest and lowest adjusted r^2 as obtained from the models are marked by blue and grey respectively in table 5.1. It was observed that model without intercept using mean radiance from gain 50 dB image outperformed all the other models. This model produced the highest adjusted r^2 values for all the metrics except for *urban population per square kilometre*. On the other hand, the model with intercept with standard deviation of radiance from gain 20 dB image exhibited the lowest r^2 values. In addition, most of the metrics did not have significant correlations at the 95% confidence interval with this model. Census variables such as *female literates per square kilometre, total workers per square kilometre, percentage of households using electricity as power source, and PCDDP* did not have significant adjusted r^2 values in the model with mean radiance from gain 20 dB image. For model using standard deviation radiance, *number of households per square kilometre, total population per square kilometre, female literates per square kilometre, total workers per square kilometre, percentage of permanent houses, percentage of households using electricity as power source and PCDDP* did not have significant model adjusted r^2 values. The adjusted r^2 values were also not significant for some models using mean and standard deviation of radiance from the gain 20 dB image. These values are highlighted in red boxes in table 5.1.

Models for *percentage of households having access to electricity* and mean and standard deviation of radiance from gain 20 dB image produced negative adjusted model r^2 values, represented as 0.00 (Zar, 2010) in the table 5.1. This was caused by the correlation coefficient of the whole population from which the sample was drawn (in this case all the 32 districts) being very near to zero (Zar, 2010). The adjusted r^2 values were higher for models with no intercepts than those with intercepts. This difference is caused by the method used for calculating the model Sum of Squares (SS) (Truong, 2007).

Although some of the linear regression models demonstrated high to moderate adjusted r^2 values, an attempt was also made to test the results from multiple regression models. These models were proposed using the information from both gain 20 dB and gain 50 dB images together. The details of these models are explained in the following section.

5.4.2 Discussion of multiple regression models

Multiple regression models were calculated with more than one independent variable. Three types of models were tested: i) Models with both mean and standard deviation of radiance from gain 20 dB and gain 50 dB images; ii) Models with mean and standard deviation of radiance from gain 20 dB image and iii) Models with mean and standard deviation of radiance from gain 50 dB image. All these models were calculated both with and without intercepts.

Multiple regression models were calculated in this study using the method of backward elimination (Zar, 2010). In this method, to begin with, all the independent variables were used to fit in regression equation. The null hypothesis was tested for each of partial correlation coefficients. When the null hypothesis was rejected, all the independent variables were considered for the model. However, if the null hypothesis was true the independent variable with the lowest $|t|$ value was eliminated from the equation and the remaining variables were considered as to fit to the multiple regression equation. The step continued until all the partial regression coefficients were significant (Zar, 2010).

Apart than this, there are several other ways of calculating multiple regression models. These include: Fitting all possible equations, Forward addition of variables and Stepwise Regression (Zar, 2010). In the method of fitting all possible equations, regression equations are calculated for all the independent variables and the equation with the highest adjusted r^2 value is chosen as the best model. However, one of the drawbacks of this model is the large number of equations to be calculated, the number being $2^m - 1$ where m is the number of independent variables (Zar, 2010). In the method of forward addition of variables, the process begins with calculating regression equation with single independent variable and gradually works up by adding all the significant independent variables. The regression equation with the highest partial regression coefficient is chosen as the most suitable one. In the process of Stepwise Regression, a step – up method is followed. An independent variable is added in each step and the significance of partial correlation coefficient is tested against the values from *t*-test[11]. The equations with the non – significant *t* values are eliminated (Zar 2010).

In this research, the backward elimination method was preferred to other methods of calculating multiple regression models because it involved less computational effort. Also in method like forward addition of variables, the inclusion of a new variable did not determine whether the previously added variables were still significant or not (Zar 2010). In this study multiple regression models were calculated with and without intercepts. The variables that were pooled out and included in each of the models are shown in table 5.2.

Table 5.2: Variables pooled and included in multiple regression models

Independent variables in the models	Variables pooled in stepwise regression	Variables included in multiple regression
A) Mean and standard deviation radiance (Gain 20 dB and Gain 50 dB images) with Intercept	Mean radiance gain 20 dB Standard deviation radiance gain 20 dB Standard deviation radiance gain 50 dB	Mean radiance gain 50 dB
B) Mean and standard deviation radiance (Gain 20 dB and Gain 50 dB images) with No Intercept	Mean radiance gain 20 dB Standard deviation radiance gain 20 dB	Mean radiance gain 50 dB Standard deviation radiance gain 50 dB
C) Mean and standard deviation radiance (Gain 20 dB) with No Intercept	Standard deviation gain 20 dB	Mean radiance gain 20 dB
D) Mean and standard deviation	Mean radiance gain 20 dB	none

[11] A ***t*-test** is any statistical hypothesis test in which the test statistic follows a Student's t distribution if the null hypothesis is supported. A student's t-distribution is a continuous probability distribution that arises when estimating the mean of a normally distributed population for small sample size.

radiance (Gain 20 dB) with Intercept	Standard deviation radiance gain 20 dB	
E) Mean and standard deviation radiance (Gain 50 dB) with No Intercept	none	Mean radiance gain 50 dB Standard deviation radiance gain 50 dB
F) Mean and standard deviation radiance (Gain 50 dB) with Intercept	Standard deviation radiance gain 50 dB	Mean radiance gain 50 dB

From the table 5.2 it was observed that mean radiance from gain 50 dB image was included in the models A and F. Since none of the variables were significant in model D, there was no model proposed using mean and standard deviation of radiance with intercept from gain 20 dB image. For models B and E, only mean and standard deviation of radiance from gain 50 dB image were included. From table 5.2, it was found that only one multiple regression model using mean and standard deviation of radiance from gain 50 dB image could be proposed. The adjusted r^2 values from this model for all the variables are presented in table 5.3. All the values are significant at the 95% confidence interval.

Table 5.3: Adjusted r^2 values from multiple regression model

Census Metrics	**Mean and standard deviation radiance gain 50 dB image (no intercept)**
Number of households per square kilometre	0.96
Total population per square kilometre	0.96
Female literates per square kilometre	0.94
Total workers per square kilometre	0.96
Percentage of households with cars, jeeps and vans	0.85
Percentage of households with television	0.96
Percentage of permanent houses	0.94
Percentage of households using electricity as power source	0.96
Urban population per square kilometre	0.74
PCDDP (Rs 1998 – 99)	0.94

The results from these models were validated over the withheld districts. The final models were chosen after validation. The results of model validation are presented in the following section.

5.5 Model validation

The models were validated over the eight withheld districts. The mean and standard deviation of brightness and stable lights from these districts were used to validate the models. The resultant predicted values were compared with the actual census metrics. Residuals between the predicted metrics and census recorded data were calculated using the following formula:

$$Difference(\%) = \frac{(Actual\ value\ \ from\ census - Estimated\ value)}{Actual\ Value\ \ from\ census} \times 100$$

Equation 5.2: Percentage of difference between census and predicted values

The models which most accurately predicted the census metrics within a margin of difference of |25|% were chosen. Figure 5.6 shows the percentage of districts predicted beyond the |25|% error margin by each model. The models with which the least error was noted were chosen as the preferred ones. These are marked by red arrows in figure 5.6.

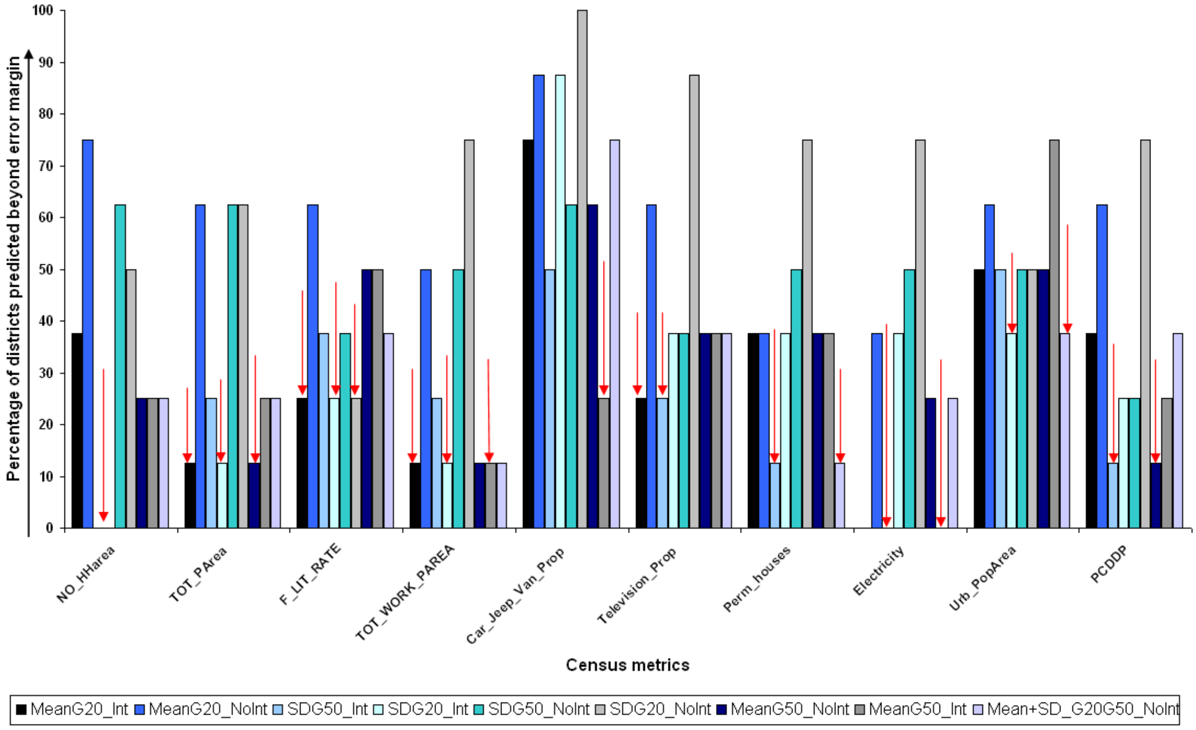


Figure 5.6: Percentage of districts with predicted values beyond |25|% error margin for each model. The red arrows indicate the best performing models.

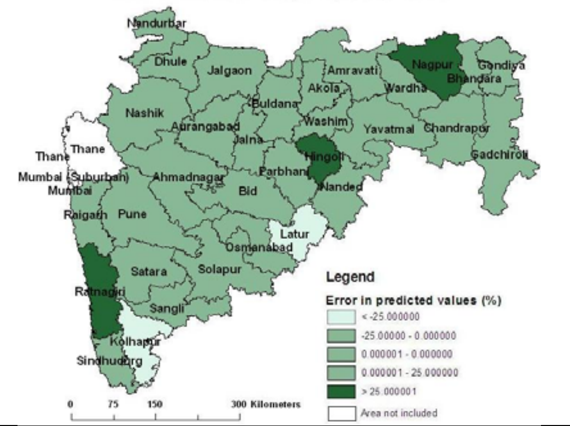
For most of the variables there was more than one optimum model. For example, for predicting *number of households per square kilometre, percentage of households with cars, jeeps and vans and PCDDP*, there were two optimum models while for *Total workers per square kilometre* there were five optimum models. In these cases the models with the highest adjusted r^2 values were chosen. The percentages of error in the predicted values were mapped (fig 5.7) in order to validate the models and determine the optimum gain.

More than 75% of the districts were predicted within the 25% error margin by the models. Variables such as *number of households per square kilometre* and number of *total workers per square kilometre* were predicted optimally for 84% of the districts. Population density, number of households with access to electricity, *PCDDP* and *percentage of households with television* were predicted for more than 75% of the districts. However, *female literates per square kilometre* and *percentage of households with cars, jeeps and vans* were optimally predicted for 41% and 47% of the districts respectively.

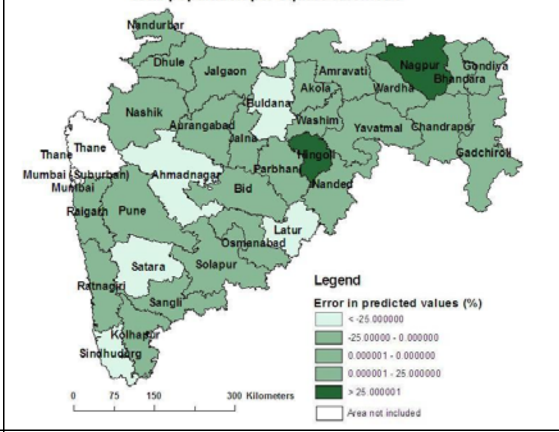
Number of households per square kilometre was predicted within 25% error for most of the districts except Ratnagiri, Kolhapur, Latur, Hingoli and Nagpur. Of these, Ratnagiri, Hingoli and Nagpur were underestimated (predicted values less than actual value in the census) while Kolhapur and Latur were overestimated (predicted values more than actual value in the census) by the models. *Population density* (*total population per square kilometre*) was overestimated for Ahmadnagar, Satara, Sindhudurg, Buldana and Latur and underestimated for Hingoli and Nagpur. *Total workers per square kilometre* was overestimated for Jalgaon, Solapur and Yavatmal while for Jalna and Gadchiroli it was underestimated. Of the demographic metrics, the highest number of errors was found in predicted values of *female literates per square kilometre*. The chosen model overestimated *female literates per square kilometre* for seven districts while it was underestimated for 12 districts. For *urban population per square kilometre*, 14 districts were overestimated while two districts were underestimated using the model.

Errors in the predicted values were also calculated for the amenities datasets. For *percentage of households with cars, jeeps and vans*, 11 districts were overestimated while 6 were underestimated. *Percentage of households using electricity as power source* was overestimated for Nandurbar, Raigarh, Jalgaon and Latur while it was underestimated for Jalna, Satara, Osmanabad and Gadchiroli. *Percentage of households with access to television* was overestimated for Raigarh, Ahmadnagar, Osmanabad, Jalgaon, Yavatmal and Nagpur while for Sangli and Latur the chosen model predicted values with more than 25% error margin. For *percentage of permanent census houses*, Pune, Yavatmal and Gondiya had more than 25% error margin while Raigarh, Jalna, Hingoli and Chandrapur had their values overestimated. *PCDDP* was overestimated for Raigarh, Solapur and Chandrapur while it was underestimated for Satara, Jalna, Amravati, Yavatmal and Gadchiroli.

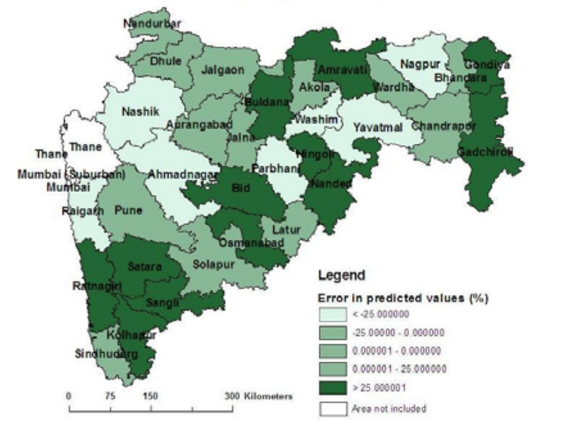
Number of households per square Kilometre



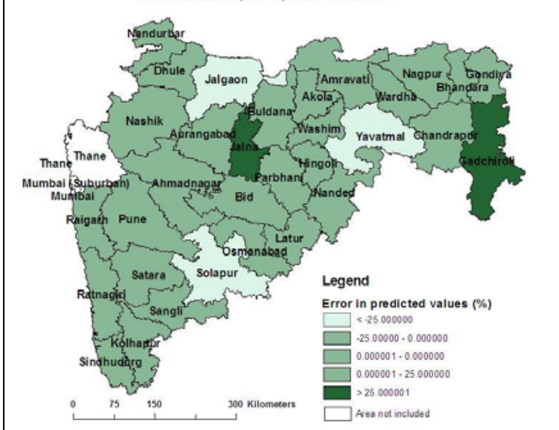
Total population per square Kilometre



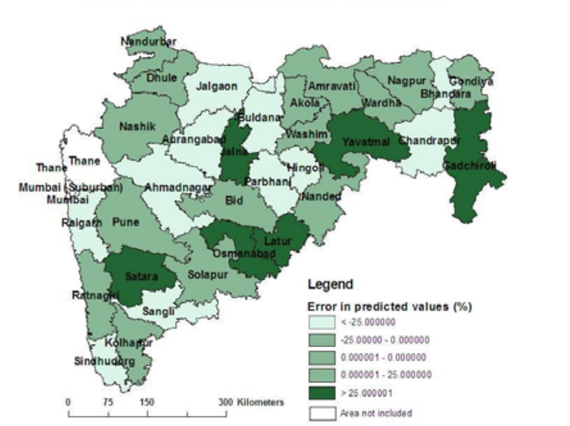
Female literates per square Kilometre



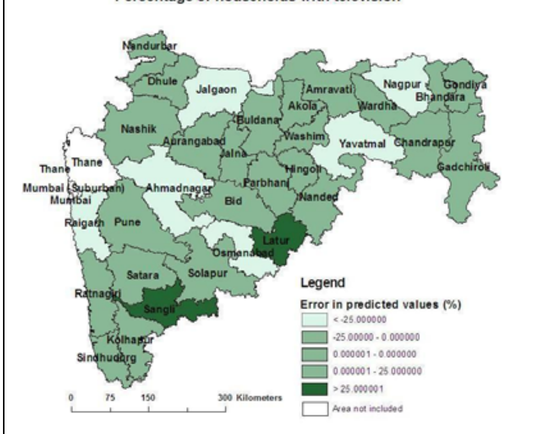
Total workers per square Kilometre



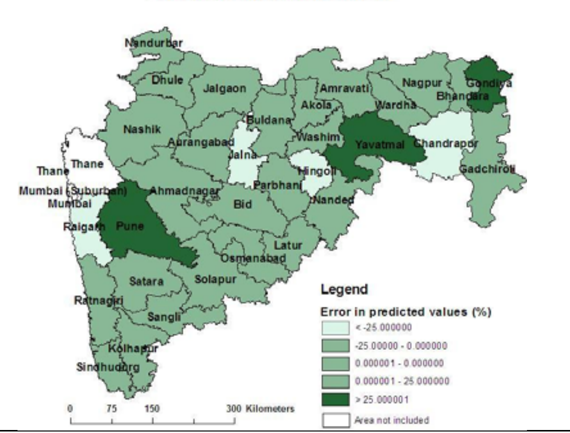
Percentage of households with cars, jeeps and vans



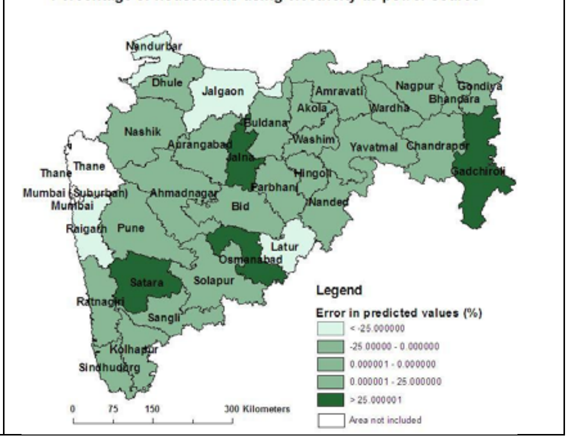
Percentage of households with television



Percentage of permanent houses



Percentage of households using electricity as power source



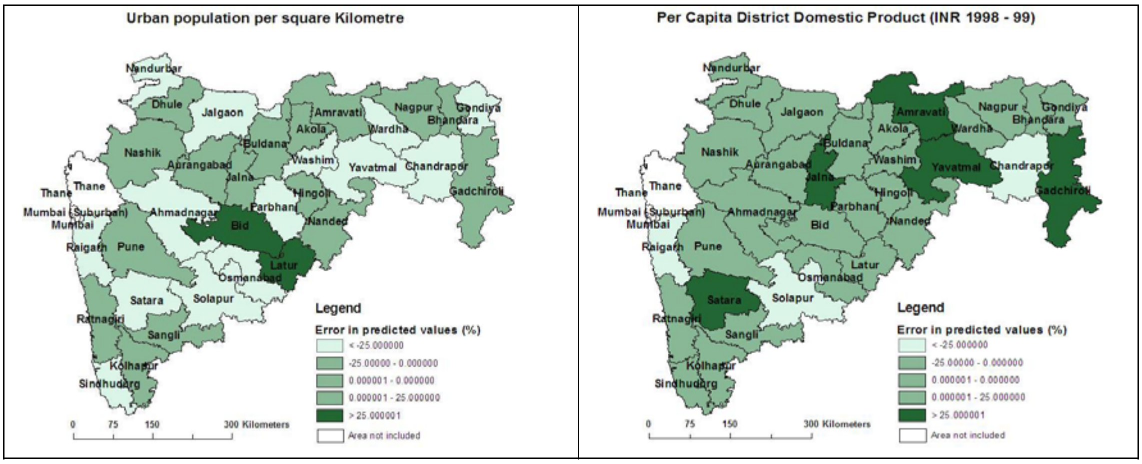


Figure 5.7: Maps showing percentages of error in the predicted values from the models for the districts

5.6 Selected models and mapping of census metrics:

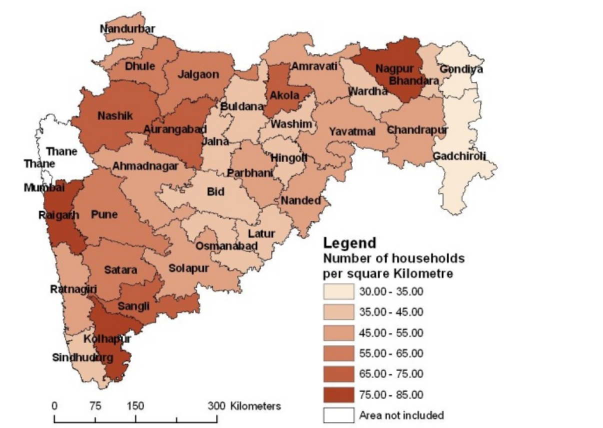
From the correlations (section 5.3) and results of model validation (section 5.5), the most suitable models were proposed for the ten metrics. The proposed models from the fixed gain images at the district level are shown in table 5.4.

Table 5.4: Selected Models

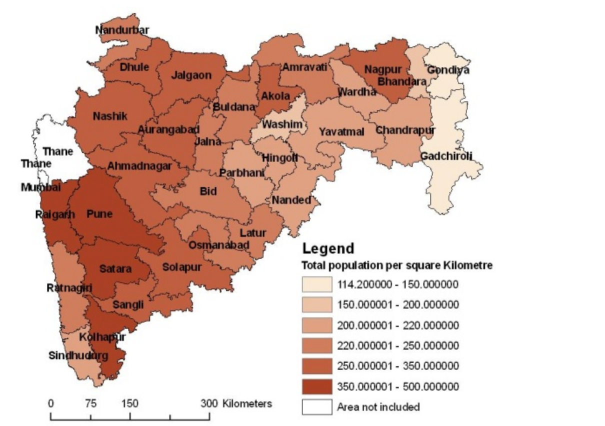
Census Metrics	Chosen Models
Number of households per square kilometre	25.06 + 30.07 x ‘standard deviation radiance from gain 50 image’
Total population per square kilometre	250.05 x ‘mean radiance from gain 50 image’
Female literates per square kilometre	12.80 x ‘standard deviation radiance from gain 20 image’
Total workers per square kilometre	108.74 x ‘mean radiance from gain 50 image’
Percentage of households with cars, jeeps and vans	-1.72 + 3.60 x ‘mean radiance from gain 50 image’
Percentage of households with television	15.58 + 19.75 x ‘standard deviation radiance from gain 50 image’
Percentage of permanent houses	56.63 x ‘mean radiance from gain 50 image’ – 17.90 x ‘standard deviation radiance from gain 50 image’
Percentage of households using electricity as power source	89.43 x ‘mean radiance from gain 50 image’ – 25.73 x ‘standard deviation radiance from gain 50 image’
Urban population per square kilometre	1.33 x ‘mean radiance from gain 50 image’ + 87.26 x ‘standard deviation radiance from gain 50 image’
PCDDP	15798.42 x ‘mean radiance from gain 50 image’

Maps were produced at the district level for the census metrics using the proposed models (table5.4). The maps are presented in figure 5.8.

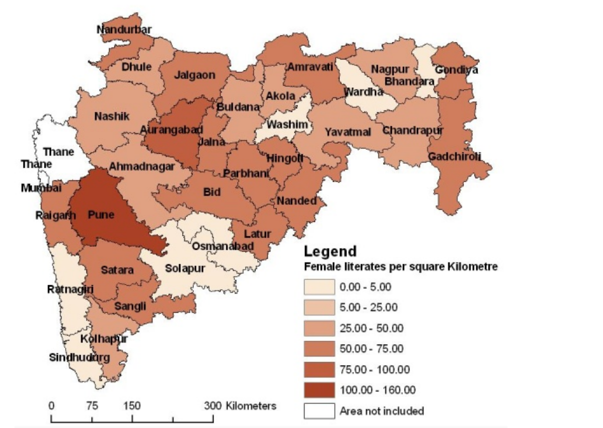
Predicted map of Number of households per square Kilometre for districts



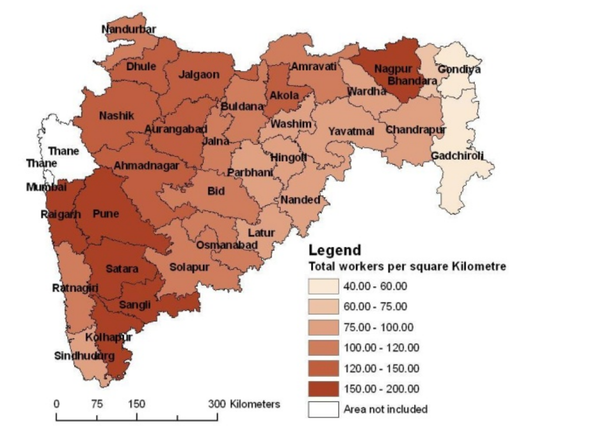
Predicted map of Total population per square Kilometre for districts



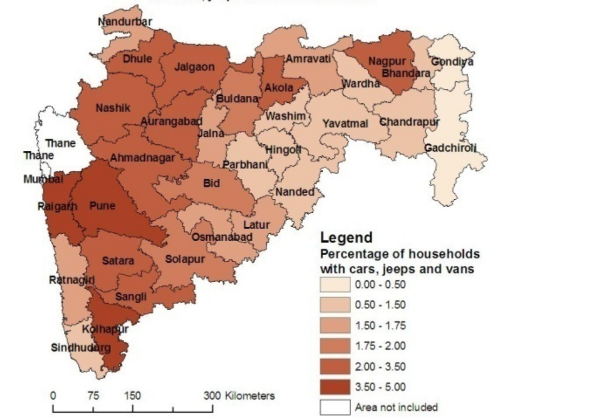
Predicted map of Female literates per square Kilometre for districts



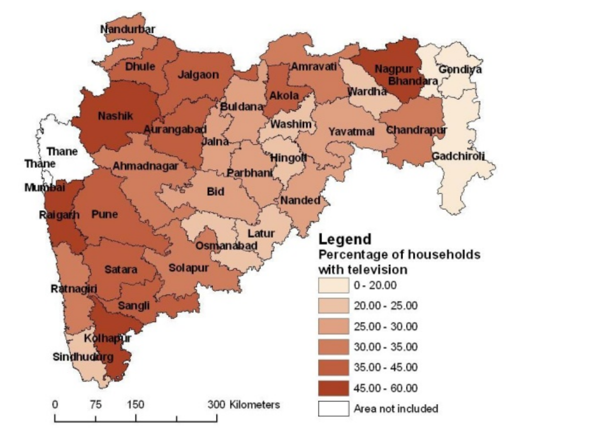
Predicted map of Total workers per square Kilometre for districts



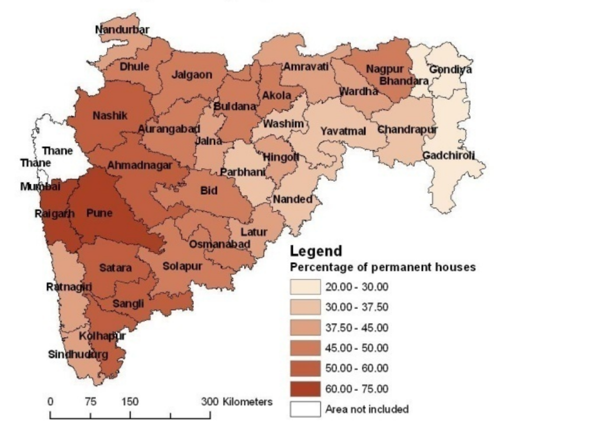
Predicted map of Percentage of households with cars, jeeps and vans for districts



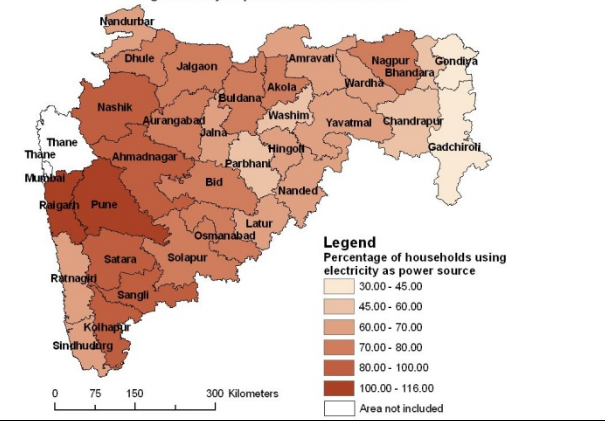
Predicted map of Percentage of households with television for districts



Predicted map of Percentage of permanent houses for districts



Predicted map of Percentage of households using electricity as power source for districts



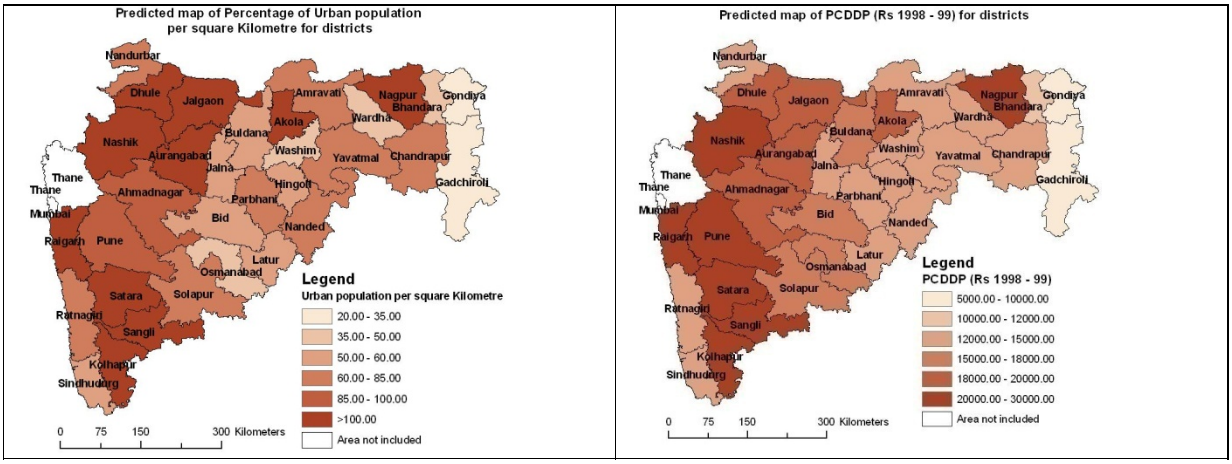


Figure 5.8: Predicted maps of census metrics for districts of the state of Maharashtra using the proposed models

The maps in figure 5.8 showed the following patterns of distribution of variables over the state of Maharashtra:

- The part of the state adjacent to MMA had higher predicted values for most of the metrics.
- The district of Raigarh had the highest predicted value for all the metrics. A part of MMA extends across the north of this district, which contributed to higher level of overall development in the area.
- Higher predicted metrics were also found southward from Raigarh through Pune, Satara, Sangli and Kolhapur.
- A separate patch of moderate to high predicted value was found in the district of Nagpur in the eastern part of the state. The city of Nagpur is the district-headquarter and is a metropolis with a population of 2.1 million. It exerts a strong migratory pull and people from surrounding rural areas move to this district. This may play a role in the higher predicted values of the metrics.
- The districts such as Yavatmal, Chandrapur, Nanded, Washim and Hingoli had lower predicted values compared to other districts in the state. These districts were mainly rural in nature and therefore, have low values for both demographic and amenities datasets.

On the basis of the results obtained from predicted maps, the districts of Maharashtra were divided into two categories:

i. Districts for which the census metrics were over-predicted by DMSP-OLS images.

ii. Districts for which the census metrics were under-predicted by DMSP-OLS images.

Table 5.5 shows the districts which were over and under-predicted in the different models.

Table 5.5: Districts over and under-predicted using the proposed models

Census metrics predicted by models	**Over-predicted (estimated value more than 25% of census recorded)**	**Under-predicted (estimated value less than 25% of census recorded)**	**Percentage of districts for which the models worked correctly (within 25% error margin)**
Number of households per square kilometre	Kolhapur, Latur	Ratnagiri, Hingoli, Nagpur	84.38
Total population per square kilometre	Buldana, Ahmadnagar, Satara, Sindhudurg, Latur	Hingoli, Nagpur	78.13
Female literates per square kilometre	Nashik, Ahmadnagar, Raigarh, Parbhani, Washim, Yavatmal, Nagpur	Amravati, Buldana, Ratnagiri, Satara, Sangli, Kolhapur, Bid, Osmanabad, Hingoli, Nanded, Gondiya, Gadchiroli	40.63
Total workers per square kilometre	Jalgaon, Solapur, Yavatmal	Jalna, Gadchiroli	84.38
Percentage of households with cars, jeeps and vans	Jalgaon, Ahmadnagar, Raigarh, Sindhudurg, Sangli, Aurangabad, Parbhani, Hingoli, Chandrapur, Bhandara	Satara, Jalna, Osmanabad, Latur, Yavatmal, Gadchiroli	46.88
Percentage of households with television	Jalgaon, Ahmadnagar, Raigarh, Osmanabad, Yavatmal, Nagpur	Sangli, Latur	75.00
Percentage of permanent houses	Raigarh, Jalna, Chandrapur	Pune, Yavatmal, Gondiya	78.13
Percentage of households using electricity as power source	Nandurbar, Jalgaon, Raigarh, Latur	Jalna, Satara, Osmanabad, Gadchiroli	75.00
Urban population per square kilometre	Nandurbar, Raigarh, Sindhudurg, Satara, Solapur, Osmanabad, Ahmadnagar, Jalgaon, Washim, Parbhani, Yavatmal, Wardha, Chandrapur, Gondiya	Bid, Latur	50.00
PCDDP	Raigarh, Solapur, Chandrapur	Satara, Jalna, Amravati, Yavatmal, Gadchiroli	75.00

From the table 5.5 it was found that *households per square kilometre and population density* were over-predicted for the district of Latur. *Population density* was also predicted with more than 25%

error for Buldana, Ahmadnagar, Satara and Sindhudurg. These districts are less densely populated and DMSP-OLS predicted values which are greater than those recorded by the Indian census. The urban centres present in these districts are on average smaller in size with less population than the class I cities (population more than 1 million). On the other hand, *number of households per square kilometre and population density* were under-predicted for districts such as Hingoli and Nagpur. These districts contain large Class I cities namely Nagpur (population 2.1 million) with a high concentration of urban dwellers and some big rural settlement such as Hingoli, which act as district headquarters. They exert strong migratory pull effects and people from surrounding rural areas move to these places. These newly arrived people contribute to the *population density* being higher than that predicted by the DMSP-OLS data. This dataset predicted a *population density* of the district based on the radiance captured and total area of the districts. In the absence of a primate structure of Indian urbanization where one big city acts as the main centre, these districts are presumed to have a lesser population per unit area, as predicted using the DMSP-OLS data.

The metrics such as *percentage of households with cars, jeeps and vans, percentage of households with television, proportion of permanent census houses* and *percentage of households using electricity as power source* were over-predicted for the district of Raigarh. Models using DMSP-OLS dataset predicted these values to be higher than those recorded by the census. This district also exhibited *PCDDP* with more than 25% difference with the actual data reported by census 2001. *Percentage of households with cars, jeeps and vans* and *percentage of households with access to electricity* were under-predicted for the districts of Satara and Jalna. *PCDDP* was also predicted with a difference of more than 25% for these districts.

There was an assumption in this study that the night-time satellite imagery represented the residential population. Therefore, the census data was used for analysis. Mumbai, Greater Mumbai and Thane constitute the Mumbai Metropolitan Area and are the most urbanized parts of the state with 100% of its population considered as urban. This area had a concentrated higher radiance value and tended to give biased results in the correlations carried out at the district level with the gain 50 dB image. Hence they were not considered for the models at the district level (marked as "area not included" in the maps in figure 5.6 and figure 5.7).

5.7 Determination of Ideal Gain

Irrespective of the gain setting of the DMSP-OLS fixed gain radiance calibrated dataset, there was significant correlation with the census metrics. However, to optimize the correlation, it was necessary to experiment with the setting of the gain of dataset image to be used. In an experiment carried out by Elvidge et al. (Elvidge et al., 1999) at the NGDC/NOAA to calibrate the radiance of the DMSP-OLS night-time dataset, gains were alternated between 24 dB and 40 dB when collecting the data. The images obtained were processed and composited into global radiance calibrated dataset for the year 1996-97. In this chapter correlation coefficients and models were used to determine the appropriate gain from non composited fixed gain radiance calibrated datasets.

The range of radiances captured by the gain 50 dB image (saturation radiance 1.52E-08 watts/cm2/sr/μm) was consistently lower than the corresponding gain 20 dB image (saturation radiance 4.80E-07 watts/cm2/sr/μm). The gain 50 dB image was more sensitive to low radiances and characterized smaller populated settlements more effectively than the gain 20 dB image. On comparing the gain 50 dB and the gain 20 dB images for the state of Maharashtra, it was apparent that the gain 20 dB appeared dark for most part of the state whilst, the gain 50 dB could discriminate some data from these areas. These differences between the two images are shown in figure 1 which illustrates the nature of the gain 20 dB and the gain 50 dB (Elvidge et al., 2007b, Doll, 2008) images more clearly. The presence of the blooming effect (Doll, 2008, Elvidge et al., 2007b) was also noted around the large coastal cities in figure 5.1. This effect was caused due to the ability of the DMSP-OLS sensor to capture diffuse light around big cities and is exaggerated where gain is set to 50 dB.

Gain 50 dB image was more suitable for predicting census metrics than gain 20 dB image on the basis of the following observations:

- The correlation coefficients between census metrics and radiance (explained in section5.4) were comparatively higher for gain 50 dB image than gain 20 dB image. The correlation coefficients ranged from 0.55 to 0.82 (at the 95% confidence interval) for gain 50 dB image while for gain 20 dB image it ranged from 0.08 to 0.53 (at the 95% confidence interval).
- All the census metrics were significantly correlated with radiance information obtained from gain 50 dB image. In contrast, most of the metrics did not exhibit statistically significant correlations with standard deviation radiance as obtained from gain 20 dB image. Also four out of ten metrics (figure 2) had no significant correlations with mean radiance from gain 20 dB image.
- Models proposed using radiance information from gain 50 dB image more accurately predicted census metrics for most of the districts (table 5) than models using radiance from gain 20 dB image.

5.8 Limitations of single orbit fixed gain datasets

The single orbit fixed gain radiance calibrated datasets were not used for further analyses in this research for predicting census metrics at small spatial scales. The data suffers from a number of problems such as from cloud cover and fires. It was also found and verified by NGDC (through personal communication) that the lights tend to vary in brightness from night-to-night (Elvidge, 2010). All three of these issues were addressed by NGDC to prepare the stable lights cloud-free global composites. However, the single night's data are overlaid with the stable lights data to analyze power outages. The stable lights data obtained from the latest version 4 data series from NGDC website (National Geophysical Data Centre, 2007b) along with global composite of brightness dataset were used further in this research to propose surrogate census for India at different spatial scales. These are explained in details in the following chapters (chapters 6 and 7).

5.9 Summary

This chapter used the single orbit fixed gain radiance calibrated DMSP-OLS datasets to propose surrogate census at district level. Although these datasets suffer from a number of shortcomings, models were proposed and validated to test utility of fixed gain datasets and also determine the optimum gain to propose surrogate census. It was found from the results that at the district level gain 50 dB image was more useful than gain 20 dB image to propose the census. The correlation coefficients were higher for gain 50 dB image than those obtained with gain 20 dB image. The values ranged from 0.55 to 0.82 (at the 95% confidence interval) for gain 50 dB image while for gain 20 dB image it ranged from 0.08 to 0.53 (at the 95% confidence interval). Overall, the results from this chapter contributed to answering research questions 2 and 3. These are: RQ2: what is the most appropriate spatial resolution or mapping unit for attributing the census metrics using DMSP-OLS night-time images? And RQ3: Which type of night-time satellite image obtained from the DMSP-OLS satellite best suits the purpose of proposing surrogate census? This chapter successfully demonstrated that comparatively higher gain datasets provide more significant results for larger regions such as districts.

Key findings from the chapter:

- Single orbit fixed gain radiance calibrated images were analyzed in this chapter.
- Gain 50 dB image was more suitable in proposing the census than gain 20 dB image.
- Comparatively higher gain datasets provide more significant results for larger regions such as districts.

6. Surrogate census from global composite images at district, taluk and village level

6.1: Introduction

This chapter uses the stable light and brightness data from the DMSP-OLS satellite to propose surrogate census for areas at different spatial scales. The chapter is divided into three parts to describe the results on the basis of the spatial scales of the areas considered. Part 6A describes the results at the district level. Part 6B deals with results at the taluk level and the results at the village level are presented in part 6C. These are the hierarchies of administrative regions in the Indian census, a village being the smallest census collection district.

For proposing the surrogate census, the global composites images were used. Both a) the stable lights and b) the radiance calibrated (brightness) data were obtained for the year 2001 (concurrent with the last complete Indian census) (fig 6.1). The detailed characteristics of these datasets are explained in section 2.1.3.

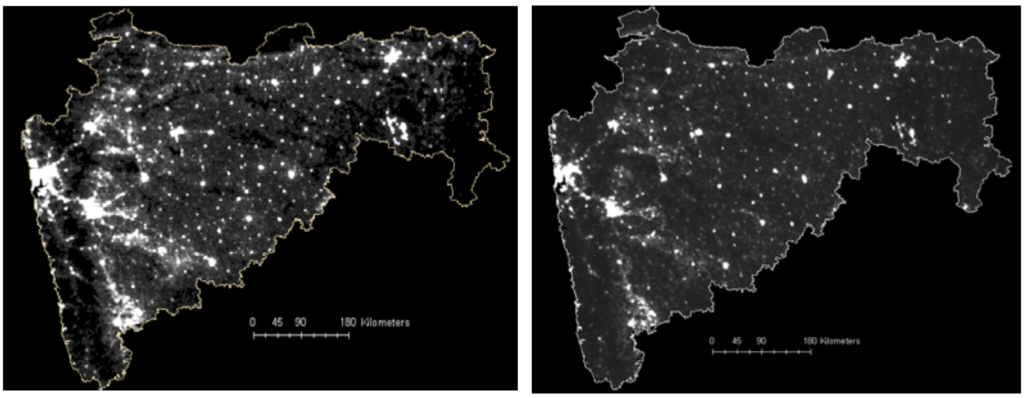


Figure 6.1: **The state of Maharashtra as obtained from two DMSP-OLS images of 2001. (a) Maharashtra shown using the stable lights dataset. (b) Maharashtra shown using the radiance calibrated dataset (showing brightness values)**

The two types of images used in the study were each processed separately. Due to the absence of on-board calibration, the stable lights product was calibrated using an empirical method (Elvidge et al., 1999). After intercalibration, the pixel values were truncated to range from 0 – 63. Stable light data obtained from F15 satellite was chosen for further analyses. Details of this method are explained in section 4.6.2. The brightness values obtained from the second DMSP-OLS image were converted from floating numbers to integers before they were used in the study. Finally, a subset of the study area was taken from both the images (fig 6.1) and the mean and standard deviation of stable lights and

brightness were calculated for the districts and used in the analyses. Details of the satellite image processing are explained in section 4.6.4. The results from both the images were then correlated with ten shortlisted metrics as obtained from the statistical tests (explained in sections 4.5 and 4.8).
Using the selected metrics and the satellite images, models were proposed for each hierarchy of administrative region. The models and their applications for each zonal scale are described in the following parts (6A, 6B and 6C).

6A Surrogate Census at District Level:

[This section has been published as the following peer reviewed publication:
Roychowdhury, K, Jones, S, Arrowsmith, C, Reinke, K & Bedford, A 2010, 'Estimating census metrics at a sub-national level using radiance calibrated DMSP-OLS night-time images', paper presented to Asia-Pacific Advanced Network (APAN) Workshop, Hanoi, Vietnam.]

The state of Maharashtra is administratively divided into 35 districts. Of these, the districts of Mumbai, Greater Mumbai and Thane were not included in the analyses to propose the models. These districts had very high values of both mean and standard deviation of brightness and stable lights compared to other districts of the state. Therefore the study considered the remaining 32 districts. From these a sample of 24 districts was randomly selected. The mean and standard deviation of brightness and stable lights from these 24 districts were correlated with the census metrics and models were proposed. The details of these steps and the results are described in the following sections.

6A.1 Discussion of correlations:

The census metrics were correlated with mean and standard deviation of stable lights and brightness as obtained from the DMSP-OLS images. The method of bootstrapping was used to calculate the correlation coefficients. The details of the process of bootstrapping are explained in section 4.8.

6A.1.1 Correlations with stable light images:

All the census metrics exhibited positive correlations with both mean and standard deviation of stable lights. The correlations were significant for all the metrics. The correlations with mean stable lights are presented in figure 6A.1 and those with standard deviation of stable lights are shown in figure 6A.2.

Moderate to strong correlations were observed between mean stable lights and the census metrics. At the 95% confidence interval, the correlation coefficients (r) with mean stable lights ranged from 0.46 for *percentage of households with access to electricity* to 0.83 for *number of persons per square kilometre.* The districts of Kolhapur, Nagpur and Pune were noted as outliers in the correlations. These three districts had high mean stable lights from the images (more than 13.5 average DN values). Some of the districts such as Gadchiroli and Sindhudurg had very low mean stable lights (less than 6.0 average DN values). These two districts also exhibited low values for most of the census metrics. The majority (18 out of 24) of the districts had average stable lights ranging from 7 – 12 DN values and were found clustered in the centre of all the correlation graphs (figure 6A.1).

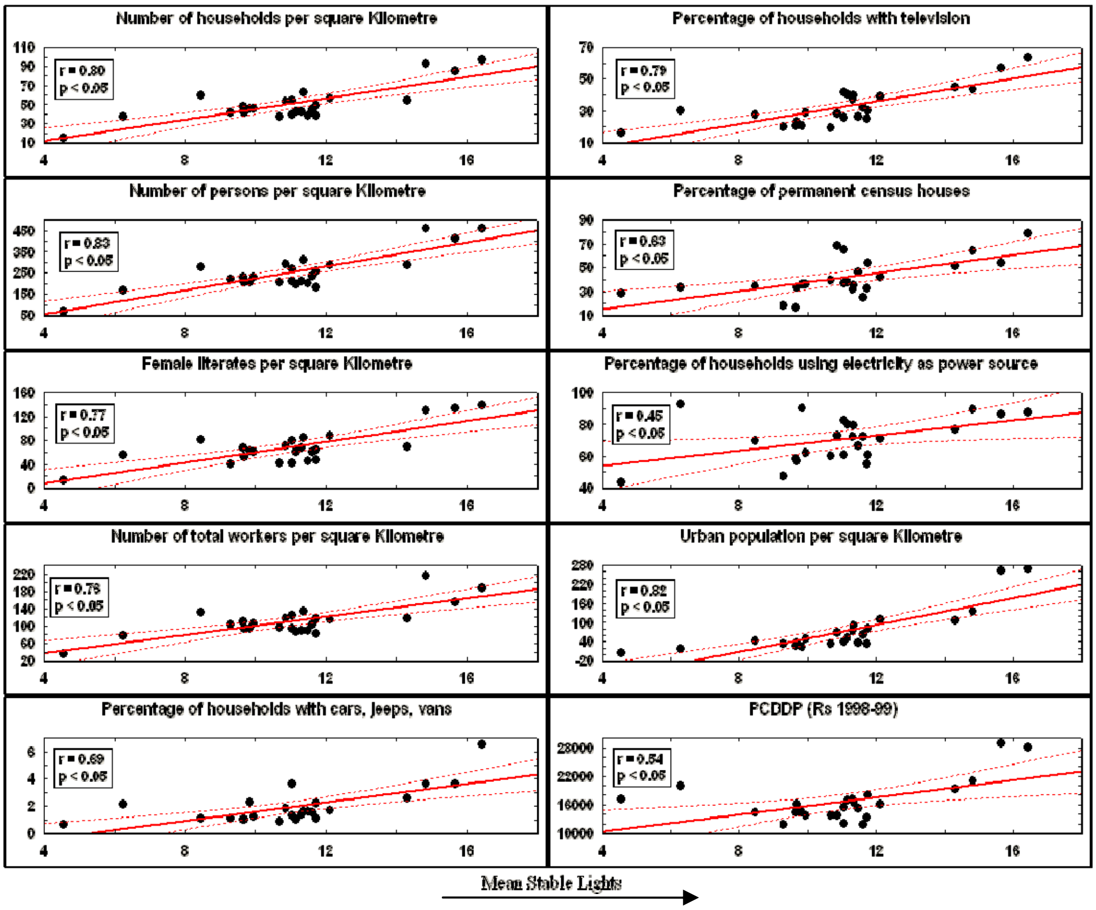


Figure 6A.1 : Correlations between census metrics and mean stable lights as recorded from global composite image of 2001

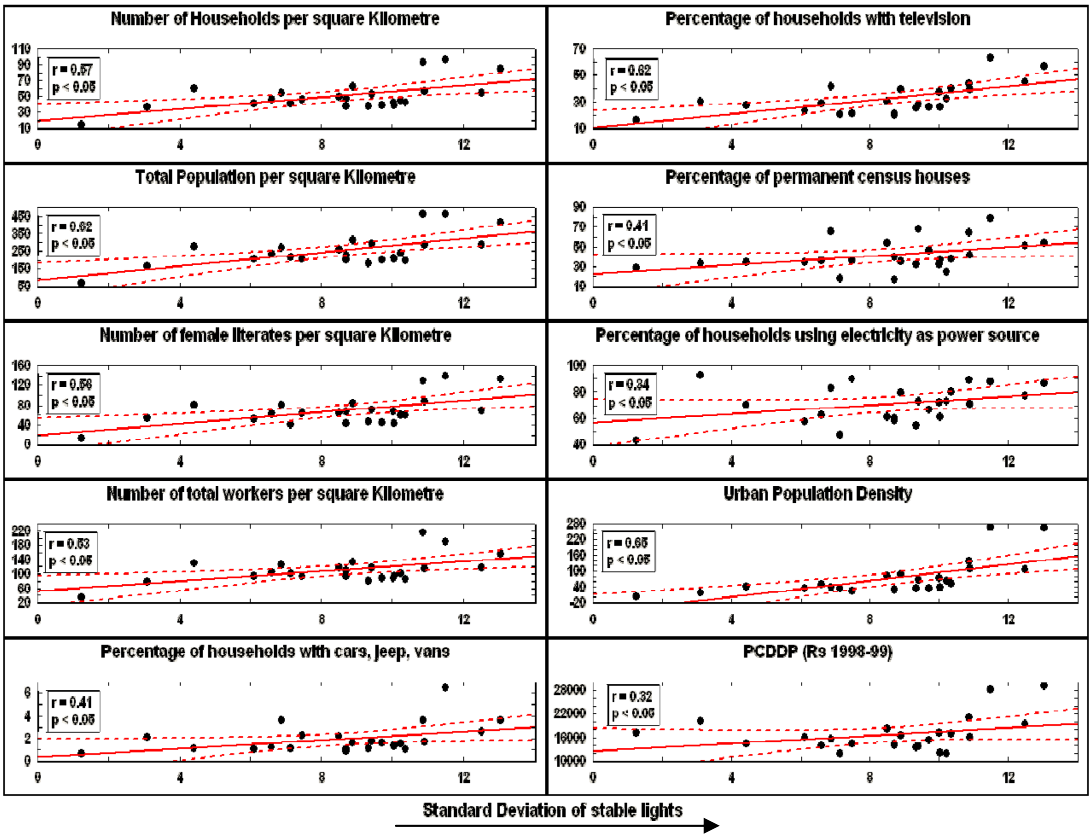


Figure 6A.2 : Correlations between census metrics and standard deviation of stable lights as recorded from global composite image of 2001

The correlation coefficients (r) obtained from correlations between census metrics and standard deviations of stable lights (figure 6A.2) were lower for all the census metrics than those obtained with mean of stable lights (figure 6A.1). The correlation coefficient (r) at the 95% confidence interval varied from 0.32 for *PCDDP* to 0.65 for *urban population per square kilometre*. There were no significant correlations between standard deviation of stable lights and *percentage of households with access to electricity* and *PCDDP* (marked by blue boxes in figure 6A.2). Districts such as Pune, Aurangabad and Nagpur, were noted as outliers in all the correlations. These three districts have class I cities (population more than 1 million) such as Pune and Nagpur as district headquarters. Due to the presence of both rural areas and large urban centres, there was high standard deviation of stable lights (more than 11 DN values) for these districts. In addition to this, because of the presence of the urban areas, these districts exhibited higher values for most of the census metrics such as *population density*, *number of households per square kilometre*, *total workers per square kilometre*, *percentage of households with access to electricity* and *percentage of households with cars, jeeps and vans*. On the other hand, for the district of Gadchiroli the absence of any major urban centres gave rise to low standard deviation of stable lights. It was plotted as one of the outliers towards the lower end of the range of standard deviation values.

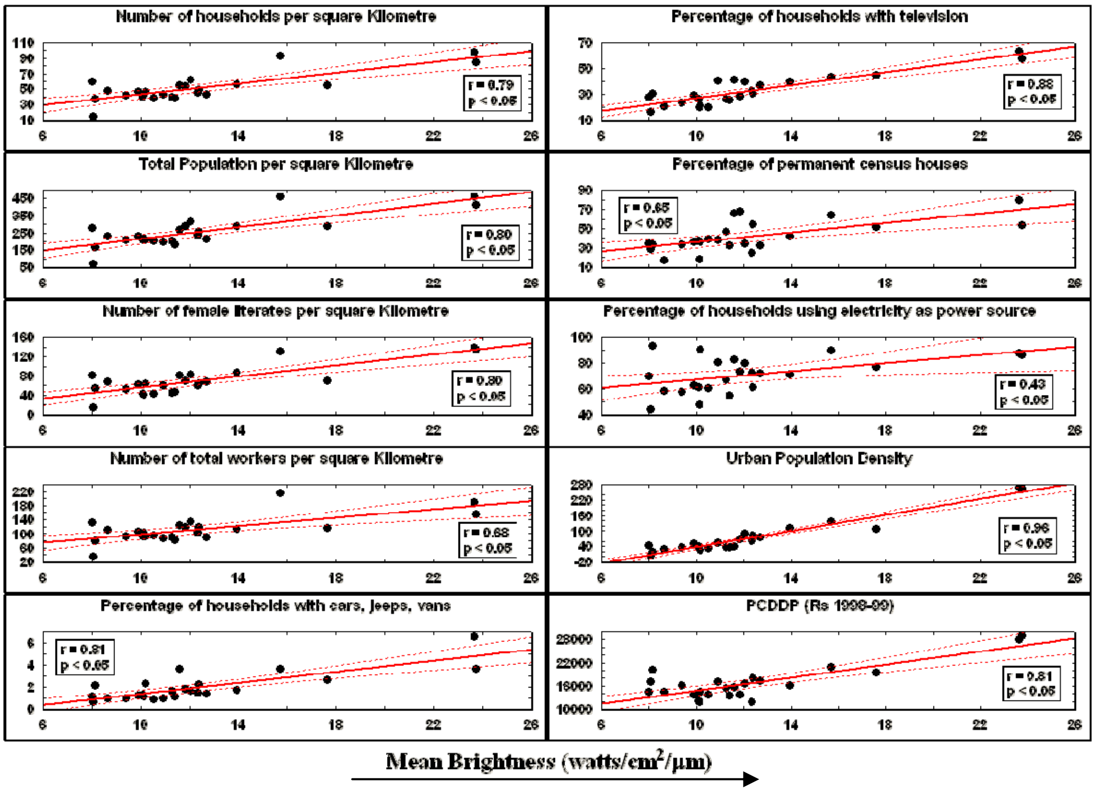


Figure 6A.3: Correlations between census metrics and mean brightness as recorded from global composite image of 2001

6A.1.2 Correlations with brightness values:

The chosen census metrics were also correlated with mean and standard deviation of brightness obtained from the global composite of brightness image. The results from these correlations are shown in figures 6A.3 and 6A.4 respectively.

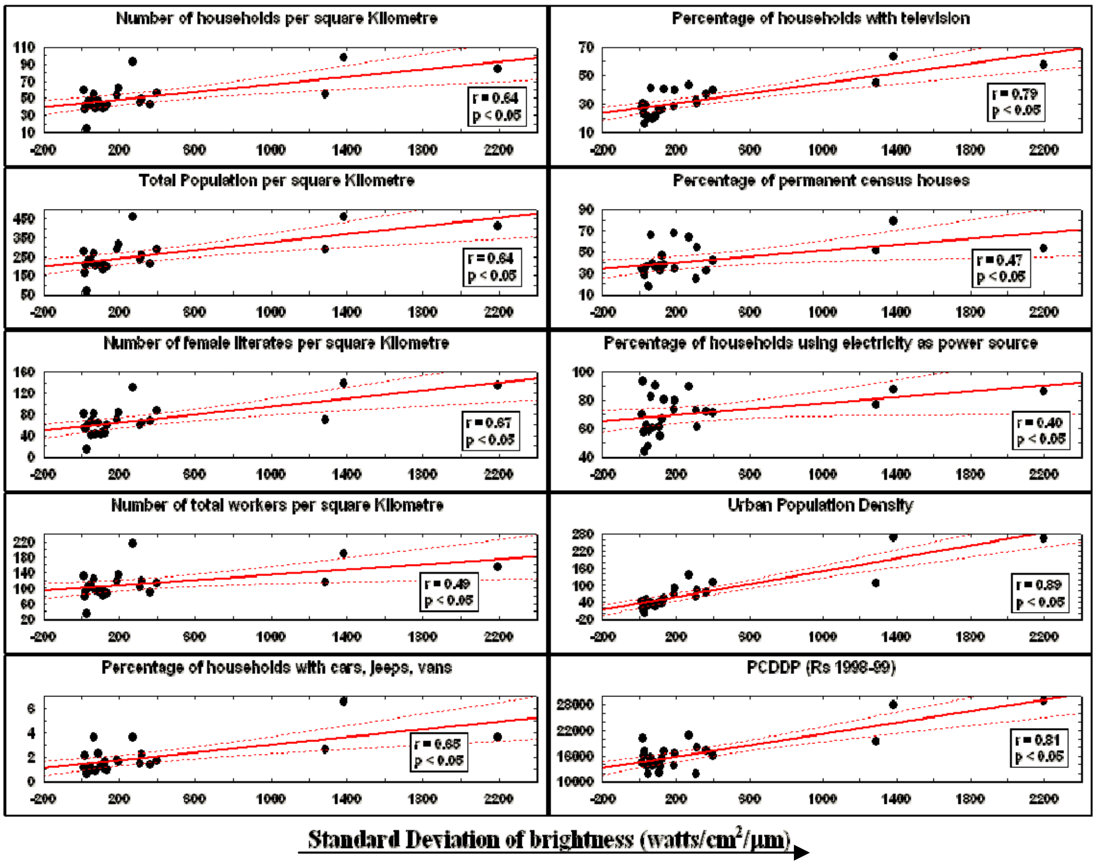


Figure 6A.4: Correlations between census metrics and standard deviation of brightness as recorded from global composite image of 2001

The correlations between mean brightness and the census metrics had higher coefficients (r) than those found with mean stable lights. The correlation coefficients ranged from 0.4 for *percentage of households with access to electricity* to 0.96 for *urban population per square kilometre*. As noted in the correlations with stable light images, the districts of Pune and Nagpur were also plotted as outliers in the correlations with mean brightness. The presence of large urban centre gave rise to more mean brightness recorded from these districts. The districts of Bhandara, Gadchiroli and Sindhudurg recorded average brightness of 8 watts/cm^2/µm each. They were plotted at the lower end of the range of mean brightness. Most of the districts demonstrated average brightness values ranging from 9 watts/cm^2/µm to 15 watts/cm^2/µm and were plotted in the centre of the graphs.

On the other hand, the correlations of census metrics with standard deviation of brightness (figure 6A.4) exhibited lower correlation coefficients (r) than those with mean brightness at the 95% confidence interval. There was no significant correlation between standard deviation brightness and *percentage of households with access to electricity* (marked in blue box in figure 6A.4).For the remaining metrics, the r values ranged from 0.49 for *number of workers per square kilometre* to 0.89 for *urban population per square kilometre*. The amenities data gave rise to higher correlations than the demographic metrics. However, in these correlations, most of the districts were clustered within standard deviation of brightness values from 50 watts/cm^2/μm to 400 watts/cm^2/μm.

6A.2 Development and discussion of the models:

6A.2.1: Discussion of linear regression models:

Simple linear regression models were calculated using the chosen metrics and mean and standard deviation of stable lights and brightness. Models with intercept and models without intercepts were tested for all the metrics. Adjusted r^2 values were calculated for each model. The equation used in the calculation of adjusted r^2 values is presented in section 5.4.1. The adjusted r^2 values from the models are presented in table 6A.1.

Table 6A.1 : Adjusted r^2 values from linear regression models. The highest adjusted r^2 values are shaded in light blue and the lowest adjusted r^2 values are shaded in grey

Census Metrics	Mean brightness with intercept	Mean brightness with no intercept	SD brightness with intercept	SD brightness with no intercept	Mean Stable Lights with intercept	Mean Stable Lights with no intercept	SD Stable Lights with intercept	SD Stable Lights with no intercept
Number of households per square kilometre	0.61	0.95	0.38	0.44	0.62	0.96	0.31	0.91
Total population per square kilometre	0.62	0.96	0.38	0.44	0.68	0.96	0.37	0.92
Female literates per square kilometre	0.62	0.94	0.42	0.48	0.57	0.93	0.29	0.89
Total workers per square kilometre	0.44	0.94	0.21	0.37	0.56	0.96	0.26	0.91
Percentage of households with cars, jeeps and vans	0.63	0.86	0.40	0.54	0.45	0.81	0.14	0.74

Percentage of households with television	0.77	0.97	0.60	0.5	0.61	0.95	0.36	0.92
Percentage of permanent houses	0.40	0.92	0.19	0.38	0.38	0.93	0.14	0.87
Percentage of households using electricity as	0.20	0.92	0.12	0.31	0.17	0.94	0.08	0.91
Urban population per square kilometre	0.92	0.81	0.78	0.8	0.66	0.71	0.40	0.68
PCDDP (Rs 1998 – 99)	0.64	0.96	0.64	0.45	0.28	0.94	0.07	0.89

The highest and lowest adjusted r^2 as obtained from the models are marked by blue and grey respectively in table 6A.1. It was observed that model without intercept using mean stable lights and brightness outperformed all the other models. These models produced the highest adjusted r^2 values for all the metrics. On the other hand, the model with intercept with standard deviation of stable lights exhibited the lowest r^2 values. Models with no intercepts had higher r^2 values compared to those with intercepts. The difference in the r^2 values were caused by the method of calculation of the model Sum of Squares (SS) (Truong, 2007). At the 95% confidence interval, the adjusted r^2 for *percentage of households using electricity as power source* were not significant for models using standard deviation of brightness with intercepts. The adjusted r^2 was also not significant at the same confidence interval for *PCDDP* for models using standard deviation stable lights with intercept. For most of the demographic metrics such as *number of households per square kilometre, total population per square kilometre, female literates per square kilometre and total workers per square kilometre* the adjusted r^2 values were higher than models using stable lights. For *total population per square kilometre*, the adjusted r^2 values were same for models using mean brightness and mean stable lights with no intercepts. For amenities datasets such as *percentage of households with cars, jeeps, vans, percentage of households with television* and *percentage of permanent census houses* the models using mean and standard deviation of brightness information exhibited higher adjusted r^2 values than the demographic datasets at the 95% confidence interval.

Multiple regression models were also tested using information from stable lights and brightness global composite images. The results from these models are detailed in the following section.

6A.2.2 Discussion of multiple regression models:

Multiple regression models with more than one independent variable were calculated. Models with three combinations of independent variables were tested: a) models with both mean and standard deviation of stable lights; b) models with both mean and standard deviation of brightness and c) models with mean and standard deviation of both stable lights and brightness information from the images. Results from these models with and without intercepts were calculated.

There are different methods of calculating multiple regression models. Of these, the method of backward elimination was used to calculate the models in this study. The details of the methods of multiple regression models are explained in section 5.4.2.

In the method of backward elimination, some variables were pooled out from the models. The variables that were pooled out and included in each multiple regression model are shown in table 6A.2.

Table 6A.2 : Variables pooled and included in multiple regression models

Independent variables in the models	**Variables pooled in stepwise regression**	**Variables included in multiple regression**
A) Mean and SD brightness and stable lights with intercept	standard deviation brightness; standard deviation stable lights	mean brightness; mean stable lights
B) Mean and SD brightness and stable lights with no intercept	standard deviation brightness	mean brightness; mean stable lights; standard deviation stable lights
C) Mean and SD of brightness with intercept	standard deviation brightness	mean brightness
D) Mean and SD of brightness with no intercept	none	mean brightness; standard deviation brightness
E) Mean and SD stable lights with intercept	none	mean stable lights; standard deviation stable lights
F) Mean and SD stable lights with no intercept	none	mean stable lights; standard deviation stable lights

Both mean and standard deviation of stable lights were included in models E and F with no variables being pooled in the process of backward elimination. In model D both mean and standard deviation of brightness were used and no variable was pooled out. In models B and C, standard deviation of brightness was pooled out while in model A, both standard deviations of brightness and stable lights were excluded and the model was proposed with only mean of brightness and stable lights.

The adjusted r^2 values of the models are presented in table 6A.3. The model adjusted r^2 values were not significant for *percentage of households with access to electricity* with models using mean and standard deviation of brightness and stable lights with intercept and mean and standard deviation of stable lights with intercept (marked in red in table 6A.3). The highest and the lowest r^2 values for each model are marked in blue and grey respectively. Overall, the r^2 values were higher than the linear regression models. The model with the mean and standard deviation of brightness and stable lights without intercept outperformed the other models. All the other models produced high to moderate adjusted r^2 values.

Table 6A.3: Adjusted r^2 values from multiple regression models. The highest adjusted r^2 values are shaded in light blue and the lowest adjusted r^2 values are shaded in grey

Census Metrics	**Mean and SD brightness and stable lights with intercept**	**Mean and SD brightness and stable lights**	**Mean and SD of brightness with intercept**	**Mean and SD of brightness**	**Mean and SD stable lights with intercept**	**Mean and SD stable lights**
Number of households per square kilometre	0.64	0.97	0.61	0.96	0.76	0.96
Total population per square kilometre	0.68	0.97	0.62	0.96	0.78	0.97
Female literates per square kilometre	0.62	0.95	0.62	0.94	0.68	0.93
Total workers per square kilometre	0.54	0.97	0.44	0.96	0.71	0.97
Percentage of households with cars, jeeps and vans	0.62	0.90	0.63	0.86	0.72	0.83

Percentage of households with television	0.76	0.97	0.77	0.97	0.65	0.95
Percentage of permanent houses	0.39	0.94	0.40	0.94	0.53	0.94
Percentage of households using electricity as power source	0.17	0.95	0.20	0.96	0.17	0.96
Urban population per square kilometre	0.92	0.88	0.92	0.92	0.71	0.68
PCDDP (Rs 1998 – 99)	0.76	0.97	0.64	0.97	0.42	0.96

6A.3 Model Validation:

Each model was validated over the withheld districts. The mean and standard deviation of stable lights and brightness from the withheld districts were used to validate the linear regression and the multiple regression models. The difference between the predicted values and values recorded by the census was calculated using the formula in section 5.5.

The models which most accurately predicted the values within a |25|% difference of error were selected. Figure 6A.5 shows the percentage of districts with predicted values beyond the |25|% error margin obtained from linear regression models and the results from multiple regression models are presented in figure 6A.6. The models which predict the metrics beyond the error margin for the most number of districts are marked with red arrows.

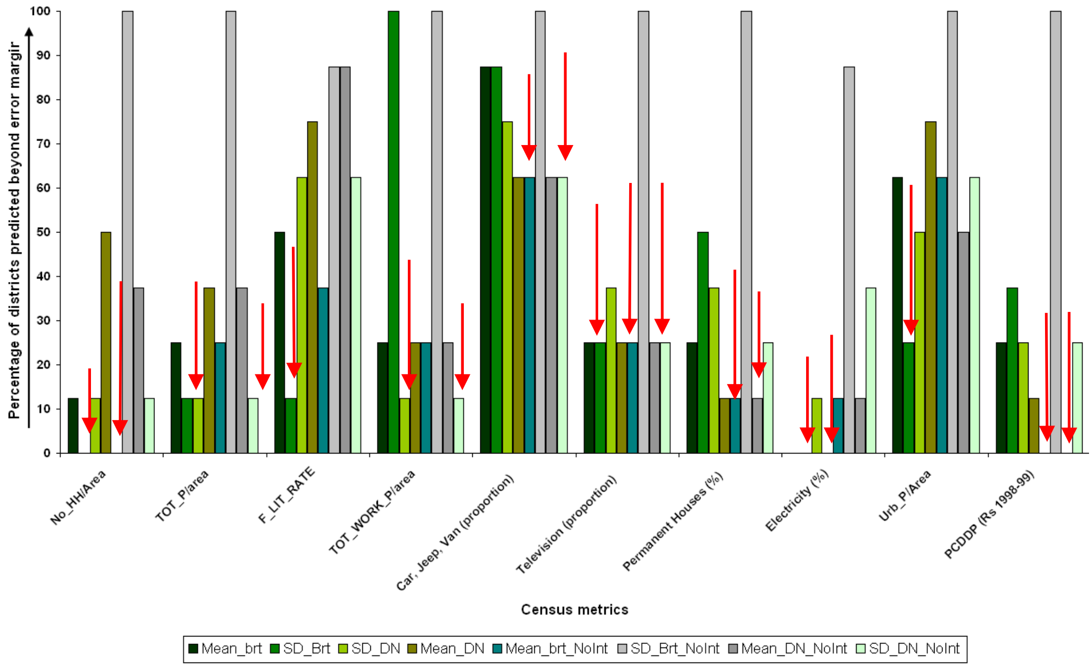


Figure 6A.5 : Percentage of districts with predicted values beyond |25|% error margin for each linear regression model. The red arrows indicate the best performing models.

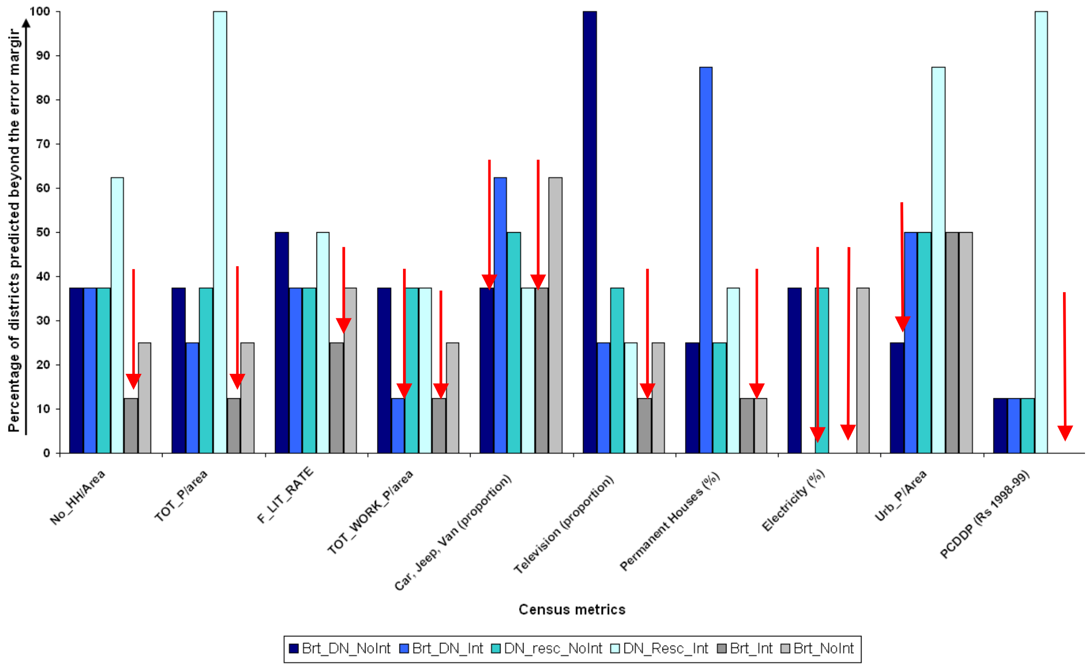
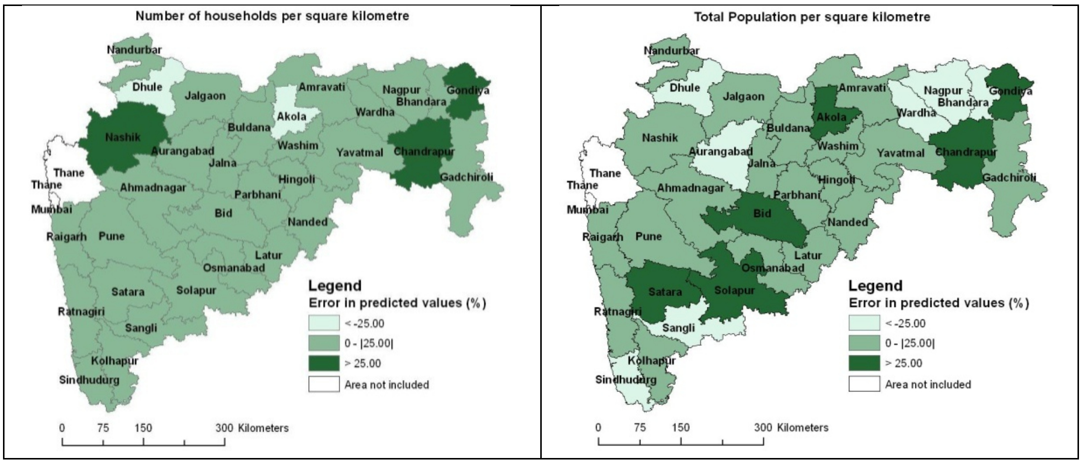


Figure 6A.6 : Percentage of districts with predicted values beyond |25|% error margin for each multiple regression model. The red arrows indicate the best performing models.

For most of the variables there was more than one optimum model. For example, there were two optimum models for *number of households per square kilometre* and *urban population per square* *kilometre*. *percentage of households with cars, jeeps and vans* was optimally predicted by three models while *PCDDP*, *total population per square kilometre* and *total workers per square kilometre* had four models predicting values with the least error. *Percentage of households using electricity as power source* was optimally predicted by seven models. In all these cases the models with the highest adjusted r^2 values were chosen. The percentages of errors in the predicted values from the selected models were mapped (figure 6A.7).

Most of the models performed with more than 70% accuracy for most of the metrics. *PCDDP* was predicted optimally for 87.5% of the districts followed by 84.4% for *number of households per square kilometre*. However, 56% of the districts were predicted within the error margin of |25|% for *percentage of households with cars, jeeps and vans*.

Number of households per square kilometre was predicted within 25% error for most of the districts except Dhule, Nashik, Akola, Gondiya and Chandrapur. Of these Nashik, Chandrapur and Gondiya were under-predicted (estimated value less than the value recorded by the census) while Dhule and Akola were overestimated (predicted value more than the census recorded one) by the models. *Total population per square kilometre* was over-predicted for Dhule, Aurangabad, Sangli, Sindhudurg, Wardha, Nagpur and Bhandara. The *population density* recorded for these districts was less than that predicted from the chosen models. On the other hand, for districts such as Satara, Solapur, Bid, Akola, Chandrapur and Gondiya, the values of *population density* was under-predicted giving rise to difference of more than +25% from the census recorded data.



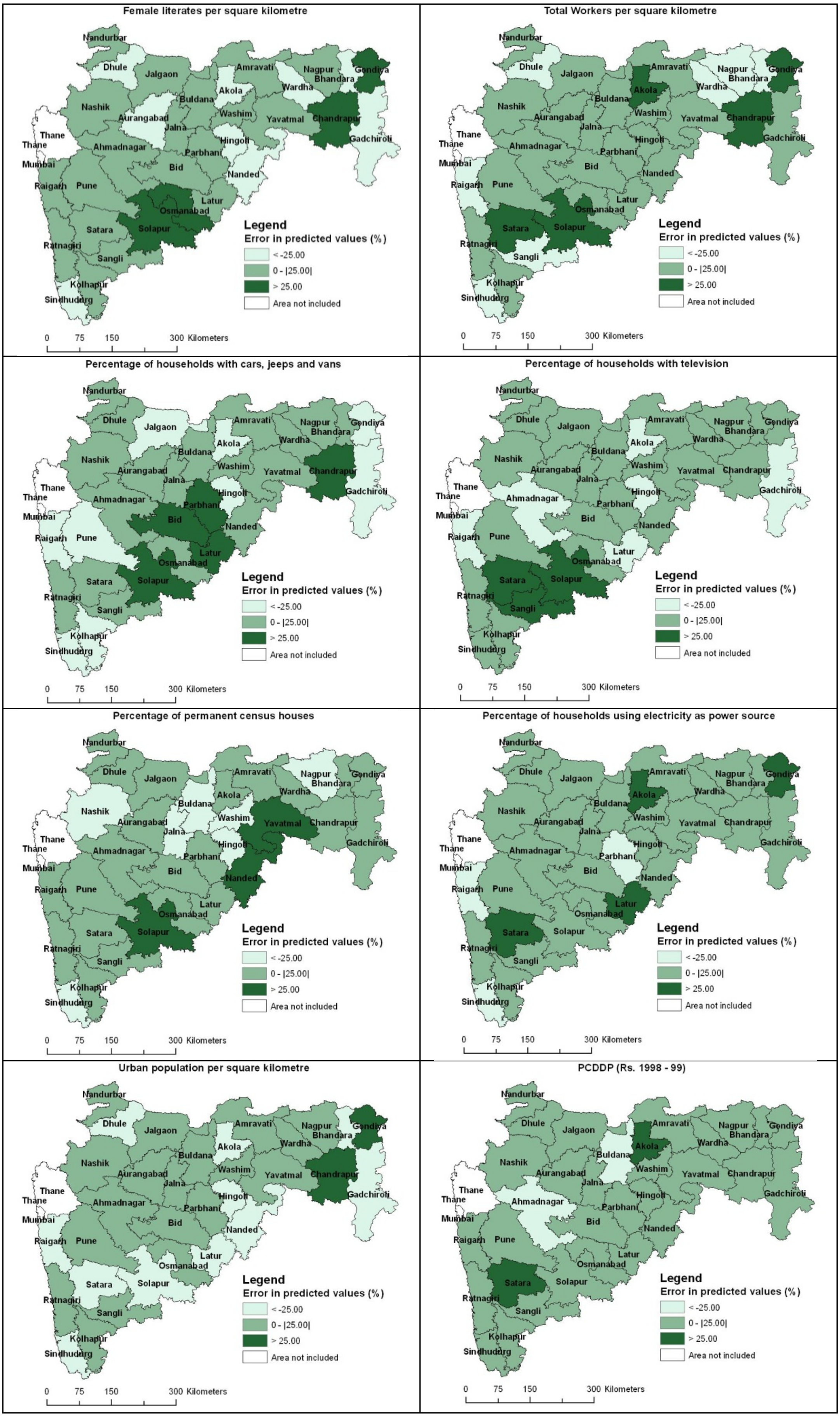


Figure 6A.7 : Maps showing percentages of error in the predicted values from the models for the districts

Female literates per square kilometre was over-predicted for Dhule, Aurangabad, Sindhudurg, Akola, Hingoli, Nanded, Wardha, Bhandara and Gadchiroli. It was under-predicted for Solapur, Osmanabad, Chandrapur and Gondiya. *Total workers per square kilometre* was predicted within the |25|% difference margin for most of the districts except for Satara, Solapur, Akola, Chandrapur, Wardha, Nagpur and Bhandara which exhibited a difference of more than a 25% between the predicted and the census recorded values. Of the demographic metrics, the highest number of districts was predicted with errors for *female literates per square kilometre* and *urban population per square kilometre*. Thirteen districts were predicted with error of more than |25|% for both the metrics. The chosen model over-estimated *female literates per square kilometre* for nine districts and under-estimated for four. *Urban population per square kilometre* was over-predicted for 11 districts and underestimated for two.

Errors were also calculated for amenities datasets. For *percentage of households with cars, jeeps and vans*, nine districts such as Pune, Raigarh, Kolhapur, Sindhudurg and Hingoli were over-estimated while it was under-predicted for five districts including Parbhani, Bid, Latur and Solapur. The least error was found for the predicted values of *PCDDP* where only two districts such as Ahmadnagar and Buldana were over-predicted and Satara and Akola were under-estimated by the chosen model. *Percentage of households using electricity as power source* was predicted within the |25|% difference margin for most of the districts except for Raigarh, Sindhudurg and Parbhani which were over-estimated and Satara, Latur, Akola and Gondiya which were under-estimated using the chosen model.

6A.4 Selected models and mapping of census metrics:

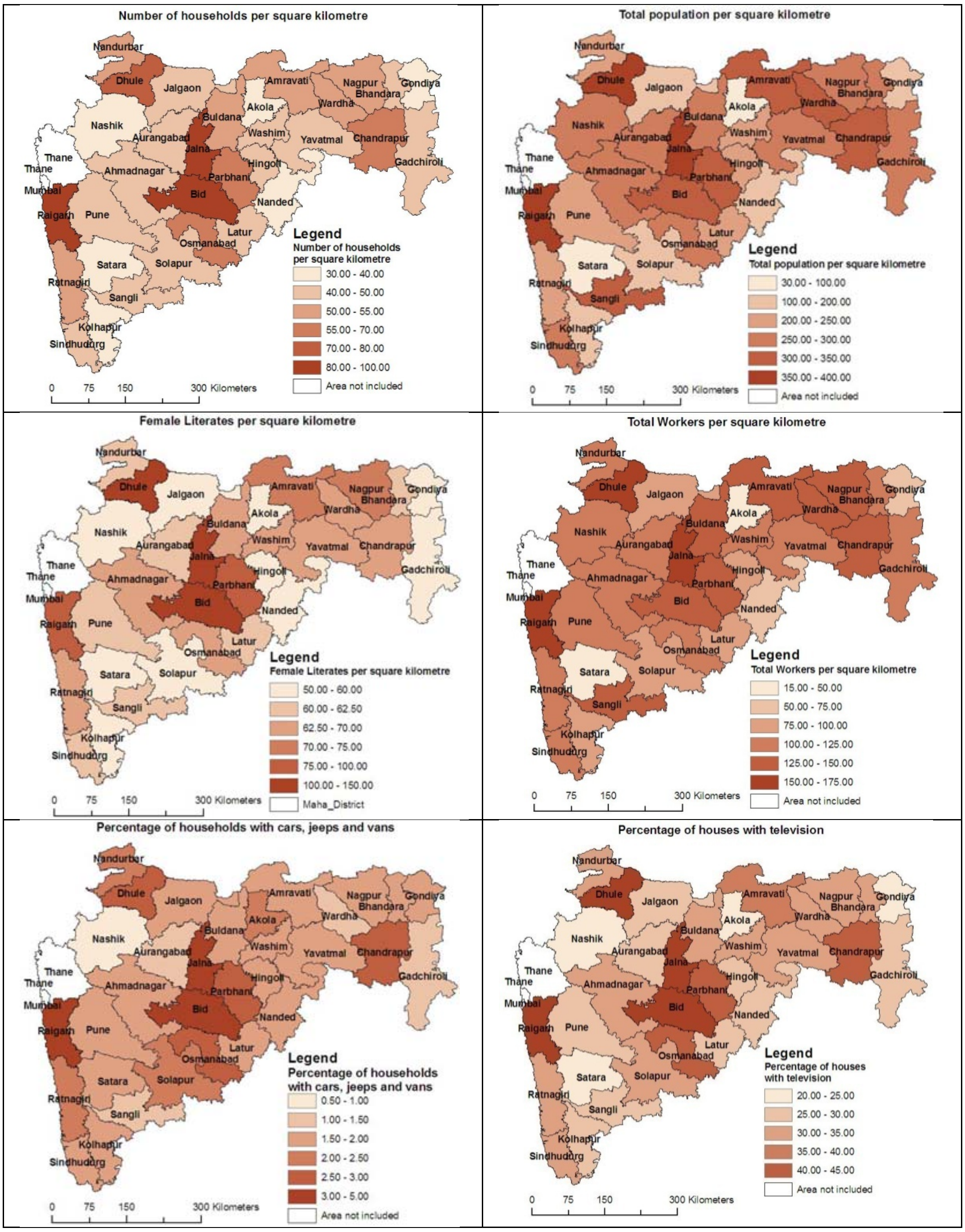
From the correlations (section 6A.2) and model validation (section 6A.3), the most suitable models for predicting the census metrics at district level were chosen. These chosen models are presented in table 6A.4.

Table 6A.4 : Optimum models

Census Metrics	Chosen Models
Number of households per square kilometre	4.08 x ‘mean brightness’
Total population per square kilometre	28.85 x ‘SD stable lights’
Female literates per square kilometre	56.64 + 0.04 x ‘SD brightness’
Total workers per square kilometre	12.48 x ‘SD stable lights’
Percentage of households with cars, jeeps and vans	0.25 x ‘mean brightness’ + 0.17 x ‘mean stable lights’ – 0.32 x ‘SD stable lights’
Percentage of households with television	1.84 + 2.50 x ‘mean brightness’
Percentage of permanent houses	3.82 x ‘mean brightness’ – 0.01 x ‘SD brightness’
Percentage of households using electricity as power source	6.85 x ‘mean stable lights’

Urban population per square kilometre	20.98 x ‘mean brightness’ – 22.04 x ‘mean stable lights’ + 4.51 x ‘SD stable lights’
PCDDP	1431.50 x ‘mean brightness’ – 3.53 x ‘SD brightness’

Maps were produced at the districts level using the chosen models (table 6A.4). These maps are presented in figure 6A.8.



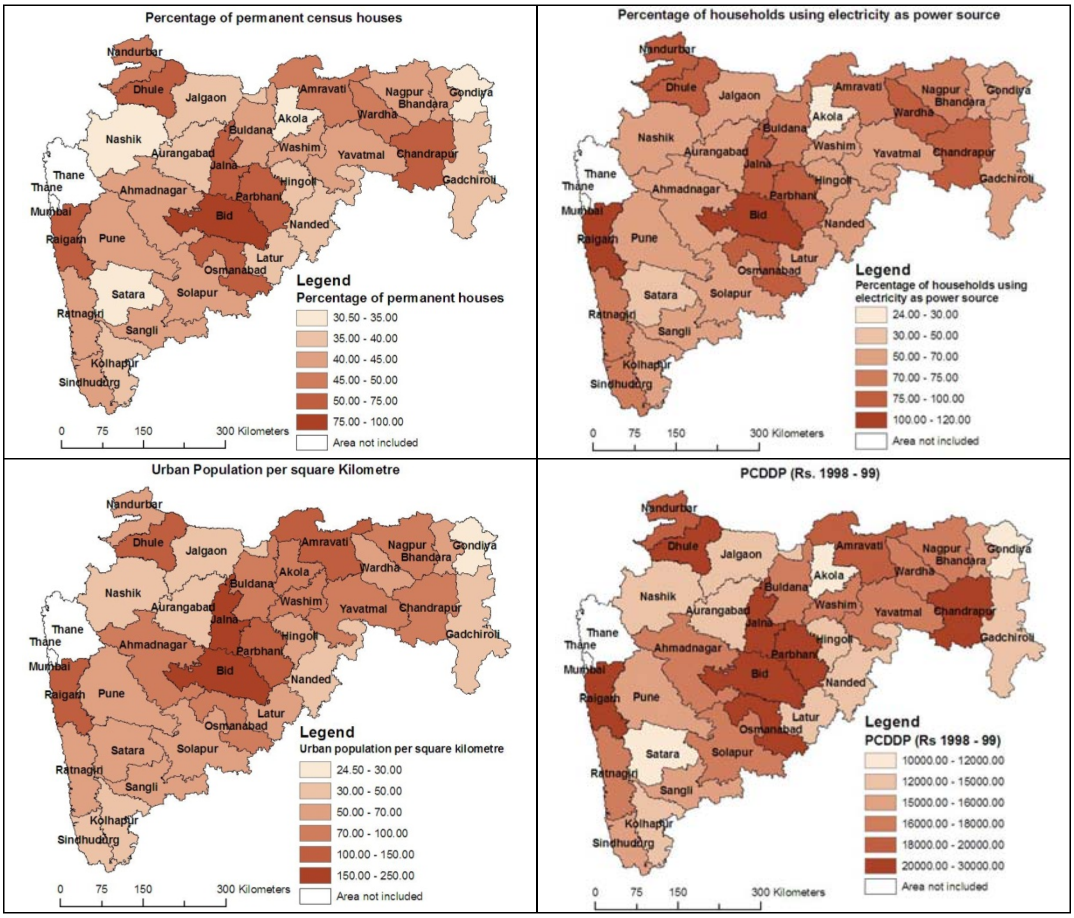


Figure 6A.8 : Predicted maps of census metrics for districts of the state of Maharashtra using the proposed models

The maps show the following patterns of distribution of census metrics across the state of Maharashtra:

- The highest predicted values for most of the metrics (except *percentage of permanent census houses* and *urban population per square kilometre*) were found in the district of Raigarh, south of MMA.
- A stretch of low predicted values for demographic metrics such as *number of households per square kilometre*, *number of female literates per square kilometre* and *urban population per square kilometre* was found along the western districts extending from Jalgaon, Nashik, Aurangabad, Ahmadnagar, Pune, Satara and Kolhapur.
- Patches of high predicted values for most of the metrics were found in the central part of the state in the districts of Bid, Jalna and Parbhani.
- Moderate to low predicted values for all the metrics were located in the districts of Gadchiroli, Gondiya and Bhandara in the east and Kolhapur and Sindhudurg in south.
- Overall, moderate to low predicted values were found in the western districts (surrounding MMA) and eastern districts while higher predicted values were recorded in the central part of the state.

On the basis of the maps illustrated in figures 6A.7 and 6A.8, the districts of Maharashtra can be divided into two categories:

i. Districts for which the census metrics were over-predicted by DMSP-OLS images;
ii. Districts for which the metrics were under-predicted by the DMSP-OLS images.

Table 6A.5 shows the districts which were over-predicted and under-predicted by the DMSP-OLS images.

Table 6A.5: Districts over and under-predicted using the proposed models

Census metrics predicted by models	**Over-predicted (estimated value more than 25% of census recorded) districts**	**Under-predicted (estimated value less than 25% of census recorded) districts**	**Percentage of districts for which the models worked correctly (within 25% error margin)**
Number of households per square kilometre	Dhule, Akola	Nashik, Chandrapur, Gondiya	84.4
Total population per square kilometre	Dhule, Aurangabad, Wardha, Nagpur, Bhandara, Sangli, Sindhudurg	Gondiya, Chandrapur, Akola, Bid, Solapur, Satara	59.4
Female literates per square kilometre	Dhule, Aurangabad, Akola, Wardha, Bhandara, Gadchiroli, Nanded, Hingoli, Sindhudurg	Gondiya, Chandrapur, Osmanabad, Solapur	59.4
Total workers per square kilometre	Dhule, Wardha, Nagpur, Bhandara, Sangli, Sindhudurg, Raigarh	Akola, Gondiya, Chandrapur, Satara, Solapur	62.5
Percentage of households with cars, jeeps and vans	Jalgaon, Pune, Raigarh, Kolhapur, Sindhudurg, Akola, Hingoli, Gondiya, Gadchiroli	Chandrapur, Parbhani, Bid, Latur, Solapur	56.3
Percentage of households with television	Raigarh, Ahmadnagar, Akola, Hingoli, Latur, Gadchiroli	Satara, Sangli, Solapur	71.9
Percentage of permanent houses	Nashik, Jalna, Buldana, Washim, Hingoli,	Solapur, Nanded, Yavatmal	68.8

	Nagpur, Sindhudurg		
Percentage of households using electricity as power source	Raigarh, Sindhudurg, Parbhani	Gondiya, Akola, Latur, Satara	78.1
Urban population per square kilometre	Dhule, Akola, Bhandara, Gadchiroli, Nanded, Hingoli, Latur, Solapur, Satara, Sindhudurg	Chandrapur, Gondiya	62.5
PCDDP	Buldana, Ahmadnagar	Akola, Satara	87.5

From table 6A.5, it was found that all the demographic metrics were over-predicted for the district of Dhule. For Akola, Wardha, Bhandara, Sangli and Sindhudurg *total population per square kilometre* and *total workers per square kilometre* were predicted with a difference of more than 25% than that recorded by the census. These districts are less densely populated with no major urban centres and DMSP-OLS predicted values were more than those recorded by the census. On the other hand for the districts of Chandrapur and Gondiya, the demographic metrics were under-predicted by the DMSP-OLS images. The district headquarters are big rural settlements and they act as major pull factors to population from surrounding areas. As a result, according to the census, these districts had higher *population density* than those predicted by the DMSP-OLS images. The models used the information obtained from lit pixels and total area of the districts. Light captured by the sensor was taken as a surrogate indicator of human habitation and therefore, the predicted values were considered as unbiased in nature.

The metrics such as *percentage of households with cars, jeeps and vans, percentage of households with television, proportion of permanent census houses* and *percentage of households using electricity as power source* were over-predicted for the district of Raigarh. This district is adjacent to MMA with extension of the urban agglomeration to the north of the district. The central and southern parts of the district were mainly darker in the images indicating areas with more rural population. There could be a disparity in the level of distribution of infrastructure over the whole district giving rise to higher predicted values than that recorded by the census. Other districts for which the amenities datasets were over-predicted were Pune, Nashik, Kolhapur, Sindhudurg and Hingoli. On the other hand, for the districts of Satara, Solapur, Sangli and Latur the amenities datasets were under-predicted with an error of more than 25% than recorded by the census. *PCDDP* was over-predicted for the districts of Buldana and Ahmadnagar while it was under-predicted for Akola and Satara.

The districts in India are divided into taluks. Information on stable lights and brightness from DMSP-OLS datasets were used to propose models for census at taluk level. The correlations and models at taluks are described in the following part of the chapter (6B).

6B. Surrogate Census at Taluk Level

[This section has been published as the following peer – reviewed publication:
Roychowdhury, K, Jones, S, Arrowsmith, C & Reinke, K 2011, 'Indian census using satellite images: Can DMSP-OLS data be used for small administrative regions?', paper presented to Joint Urban Remote Sensing Event (JURSE), 2011, 11-13 April 2011, Munich, Germany.]

According to the Indian census, there are 354 taluks in the state of Maharashtra. However, spatial data was available for only 286 taluks. Details of the data quality issues encountered at the taluks are explained in details in section 4.7. The analyses for proposing the models were conducted on the available 286 taluks. Of these, 196 taluks (75% of the total) were randomly sampled for model development and the remaining 90 taluks were withheld for model validation. Out of these 196 taluks, Pune and Thane were omitted from the model development because of very high radiance and stable lights recorded from them. The very high brightness and stable lights from these taluks tended to produce a bias in the results.

The mean and standard deviation of brightness and stable lights from the remaining 194 taluks were correlated with the census metrics. The results from these correlations are presented in the following section.

6B.1 Discussion of correlations:

The first step in the analyses at the taluk level was to test the correlations between census metrics and the mean and standard deviation of stable lights and brightness as obtained from the DMSP-OLS images. Unlike the analyses at the district level, there was no economic indicator available for the taluks. As a result, nine census metrics (five demographic metrics and four amenities data) were used for correlations and models. The nine census metrics considered are: *number of households per square kilometre, total population per square kilometre, female literates per square kilometre, total workers per square kilometre, percentage of households with cars, jeeps and vans, percentage of households with television, percentage of permanent houses, percentage of households using electricity as power source* and *urban population per square kilometre.* In addition to this, the census did not provide information on the areas of taluks. Therefore, the areas of each taluks were calculated in kilometre and the results were used for normalization of the demographic metrics.

6B.1.1 Correlations with stable light image:

All the census metrics exhibited positive correlations with both mean and standard deviation of stable lights at the the 95% confidence interval. Figures 6B.1 and 6B.2 show the correlations with mean and standard deviation of stable lights respectively.

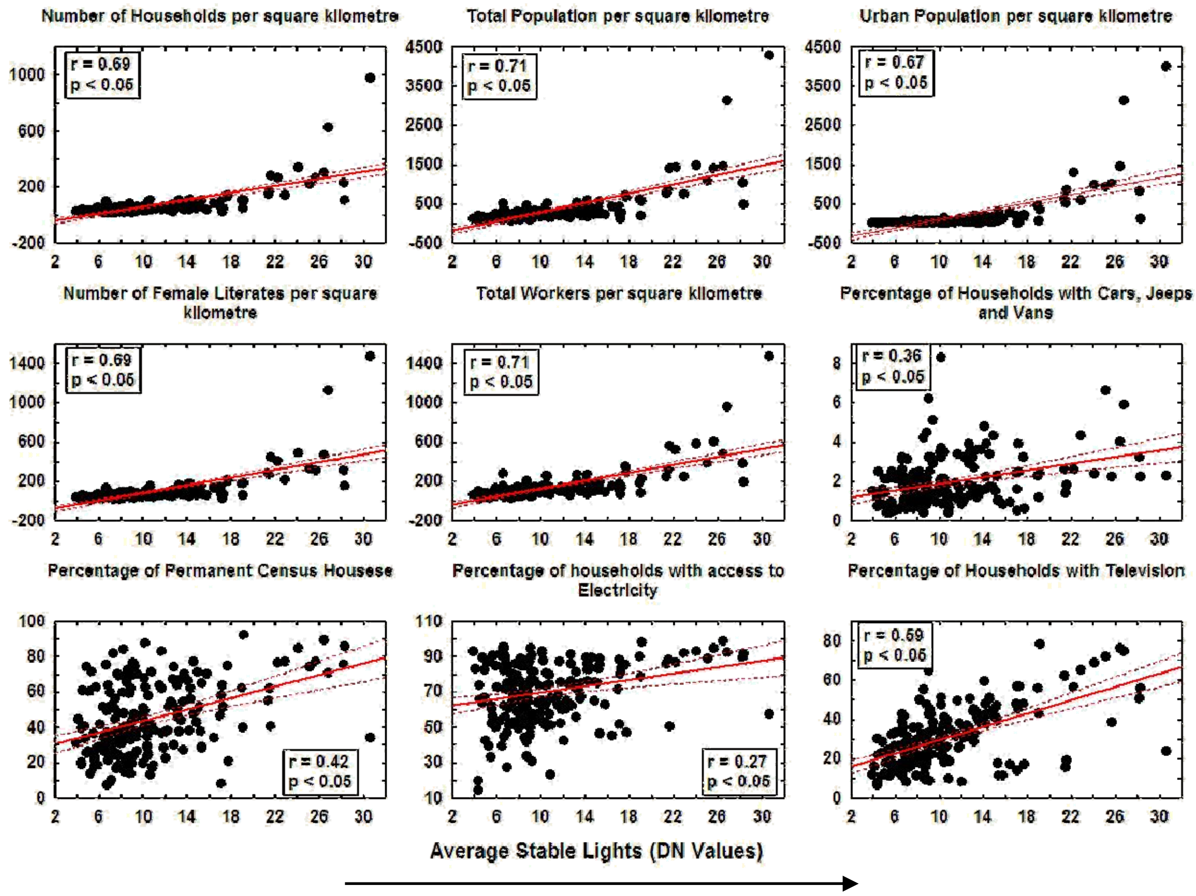


Figure 6B.1: Correlations between census metrics and mean stable lights as recorded from global composite image of 2001

Moderate correlations were noted between the census metrics and mean stable lights at the 95% confidence interval. The correlation coefficients (r) ranged from 0.36 for *percentage of households with cars, jeeps and vans* to 0.71 for *total population per square kilometre* and *total workers per square kilometre*. Most of the taluks were clustered while some taluks were noted as outliers. Around 170 taluks with average stable lights lying between 0 – 15 DN values, exhibited household density of 0 – 100 houses per square kilometre and a *population density* of 0 – 200 persons per square kilometre.

The other taluks were noted as outliers with mean stable lights ranging from more than 20 – 32 DN values. The outlier taluks are listed later in section 6B.1.3.

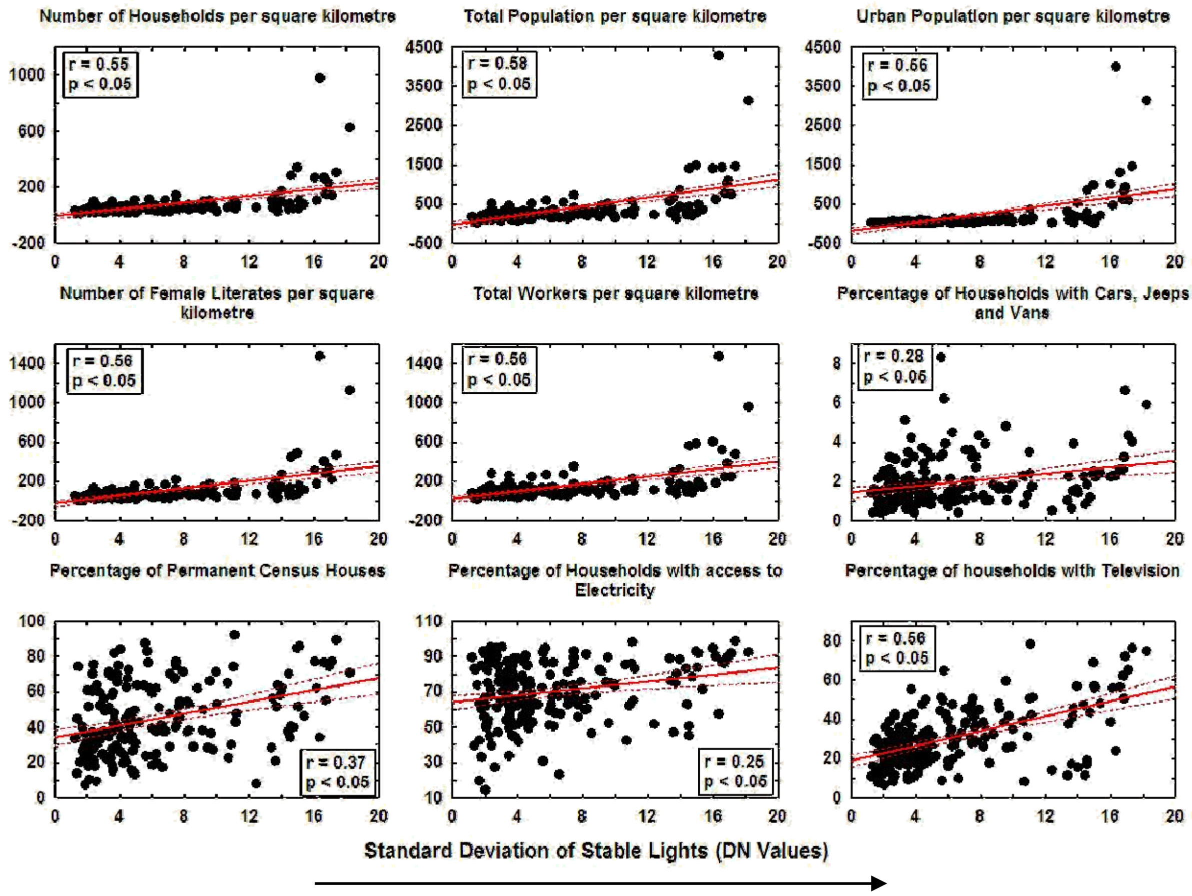


Figure 6B.2: Correlations between census metrics and standard deviation of stable lights as recorded from global composite image of 2001

When correlated with standard deviation of stable lights, the coefficients (r values) were lower than those obtained with mean stable lights. The correlation coefficient (r) ranged from 0.28 for *percentage of households with cars, jeeps and vans* to 0.58 for *total population per square kilometre*. All the correlations were calculated at the 95% confidence interval. 158 taluks exhibited standard deviation of stable lights ranging between 1 – 10 DN values. The remaining taluks had standard deviation of stable lights between 10 – 19 DN values. Some of them were also noted as outliers.

6B.1.2 Correlations with brightness image:

All the census metrics exhibited positive correlations with both mean and standard deviation of brightness at the 95% confidence interval. Figures 6B.3 and 6B.4 illustrate the correlations with mean and standard deviation of brightness respectively.

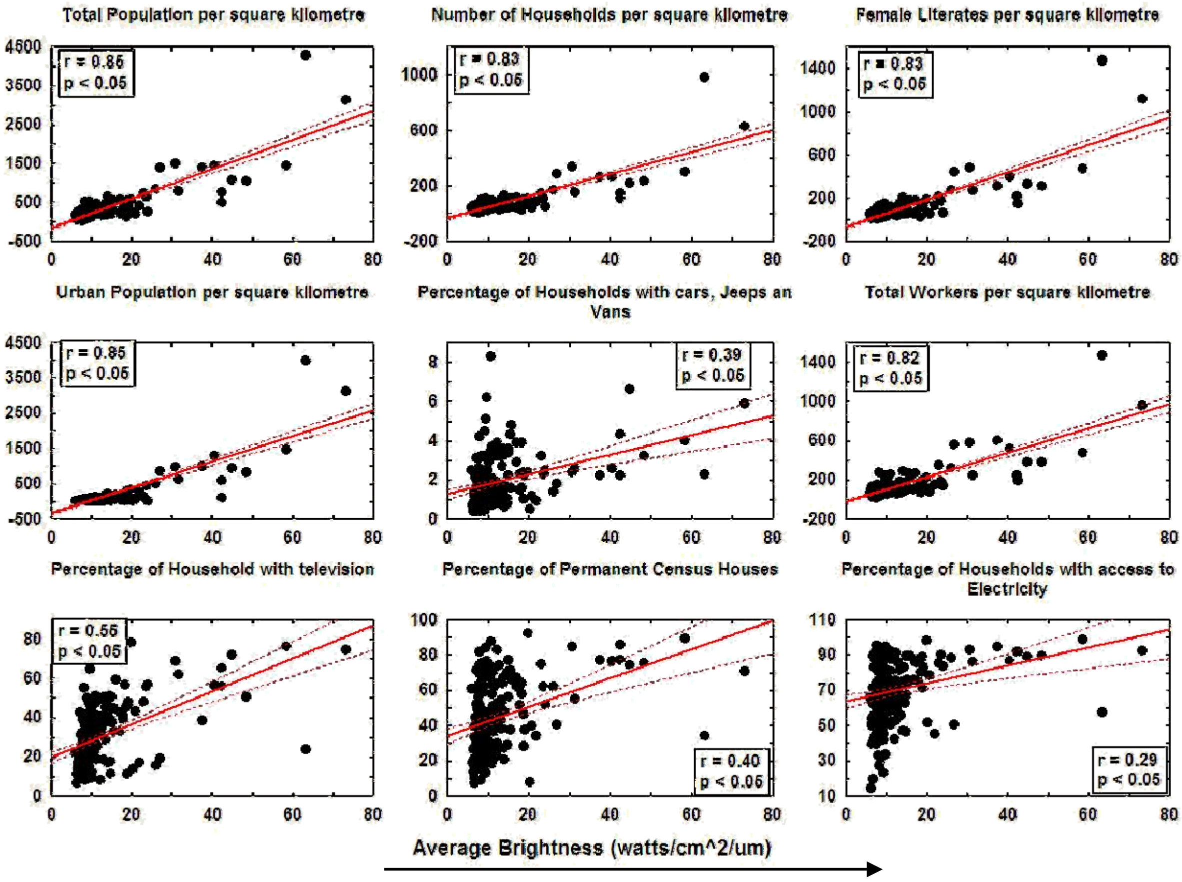


Figure 6B.3: Correlations between census metrics and mean brightness as recorded from global composite image of 2001

Moderate correlations were noted between the census metrics and mean brightness at the 95% confidence interval. The correlation coefficients (r) varied between 0.29 for *percentage of households with access to electricity* and 0.85 for *total population per square kilometre* and total *urban population per square kilometre*. Most of the taluks were clustered while some taluks were noted as outliers. For example, 175 taluks with average brightness ranging from 0 to 20 watts/cm^2/µm had household density from 0 to 200 households per square kilometre and *population density* of 0 to 500 persons per square kilometre. The other taluks were noted as outliers with mean brightness ranging from more than 22 watts/cm^2/µm to 75 watts/cm^2/µm.

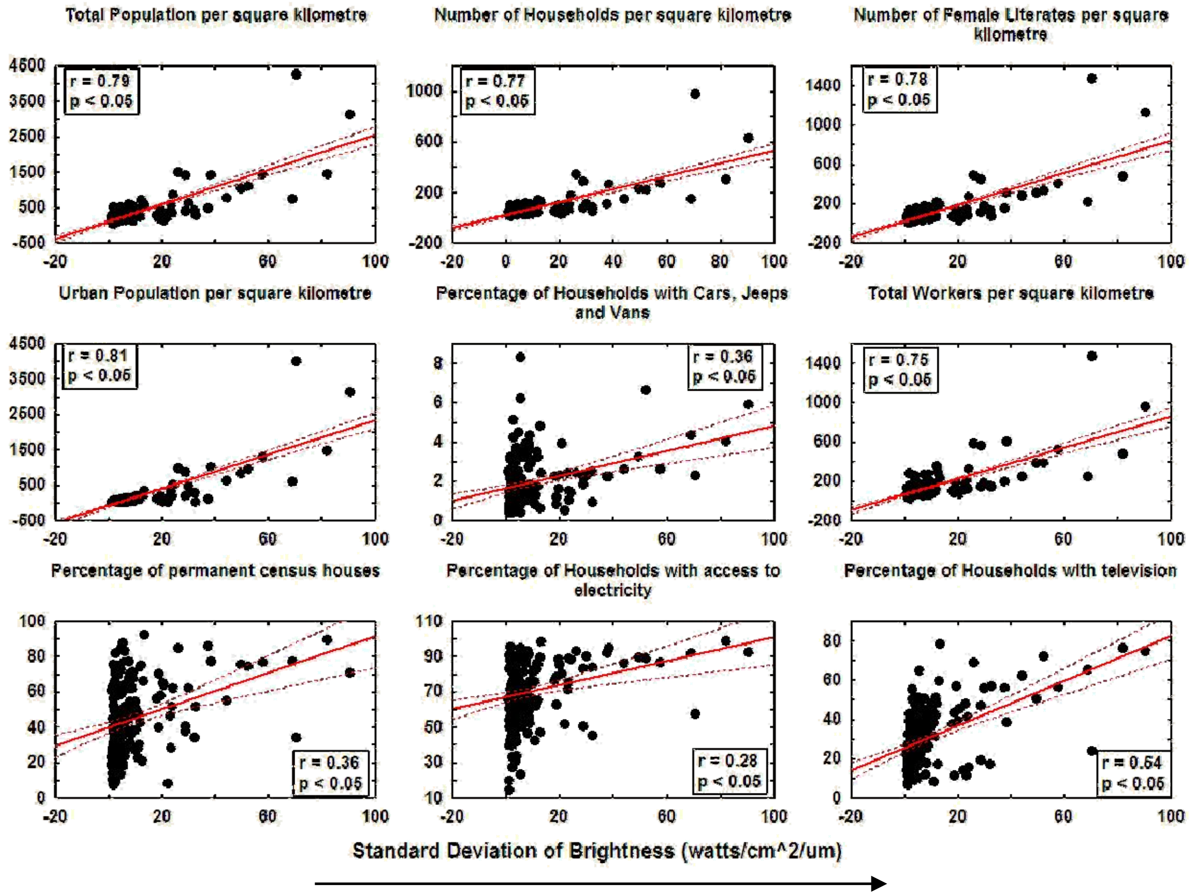


Figure 6B.4 : Correlations between census metrics and standard deviation of brightness as recorded from global composite image of 2001

When correlated with standard deviation of brightness, the coefficients (r values) were lower than those obtained with mean brightness. The correlation coefficients (r) ranged between 0.28 for *percentage of households using electricity as power source* and 0.81 for *urban population per square kilometre*. All the correlations were calculated at the 95% confidence interval. 164 taluks exhibited standard deviation of brightness ranging between 1 – 13 watts/cm^2/µm. The remaining taluks had standard deviation of brightness between 18 – 80 watts/cm^2/µm. Some of these taluks were noted as outliers.

6B.1.3 Observations from the scatter plots:

There were some common characteristics in the scatter plots obtained from the correlations between census metrics and mean and standard deviation of stable lights and brightness (figures 6B.1 to 6B.4). Both clustering and outlier taluks were noted in the correlations in all the graphs (figures 6B.1 to 6B.4).

All the graphs showed the presence of outliers. The taluks mostly noted as outliers in all the graphs are listed in table 6B.1.

Table 6B.1 : Outlier taluks

Nandgaon	Kandahar	Kalwan	Karanja
Nandgaon Khandeshwar	Vasai	Uran	Ulhasnagar
Sudhagad	Shirur	Osmanabad	Mohol
Kavathemahankal	Igatpuri	Hingna	Bhor
Bhiwapur	Baglan	Ambad	Hingoli
Kalyan	Nagpur (urban)	Kalameshwar	

The correlations with both the mean and standard deviation of stable lights and brightness exhibited the presence of outliers. It was found that the outlier taluks had very high or very low brightness and stable light values. Some of the very bright taluks included Pune, Thane, Kalyan and Ulhasnagar. Most of these taluks have no rural population and more than 90% of their households have access to electricity. On the other hand, some of the very dark outlier taluks such as Mandangad in the district of Ratnagiri has 72% of its population as casual workers and only 65% of its households use electricity as power source. There is no urban population in the taluk of Mandangad, the main centre being the village of Mandangad. The models, therefore, can be used to help distinguish between those taluks that are typically urban or rural in nature.

On the other hand, it was also observed from the scatter plots that some of the census metrics were clustered over a particular range of mean and standard deviation of brightness and stable lights. For example, 175 (out of 186) taluks with average brightness values ranging from 0 – 20 watts/cm2/μm have 0 – 200 households per square kilometre and *population density* of 0 – 500 persons per square kilometre. Around 170 (out of 186) taluks with average stable lights lying between pixel values of 0 – 15, have less than 100 houses per square kilometre and a *population density* of less than 200 persons per square kilometre. No clustering was found for *percentage of permanent houses*, *percentage of households with access to electricity* and *percentage of households with access to television* when correlated with mean and standard deviation of stable lights. However, around 123 taluks with mean and standard deviation of brightness of 0 – 10 watts/cm^2/μm were clustered within the range of 0 – 50 *percentage of permanent houses* and 40 – 95 *percentage of households with access to electricity*.

The moderate correlations between the census metrics and brightness and stable lights were caused by sampling errors. The areas of taluks ranged approximately from more than 3000 square kilometre (e.g. Etapalli) to 100 square kilometre (e.g. Pune City). As a result, the number of DMSP-OLS image pixels in these taluks, varied from 3500 to 120 (at ground sampling distance of approximately 833m). The mean and standard deviation of brightness and stable lights were calculated from these pixels. On the contrary, for larger areas such as districts, averages of more than 5000 to 20000 pixels were used for calculating the mean and standard deviation of brightness and stable lights. Therefore, more data was available for districts giving rise to higher correlation coefficients.

According to the Indian census, some of the taluks such as Jamner, Radhanagari, Bhatkuli, Goregaon, Mohol, Chincholikati, Chandvad, Vaibhawadi, Chimur, Jat, Karjat and Etapalli have no urban population. However, DMSP-OLS recorded lights over these taluks. These taluks such as Mohol (fig 6B.5), in the district of Solapur, has no urban population but there are 59 small villages located in the taluk. Of these, the villages of Mohol and Chincholikati are the main centres. These villages have 73% of their households using electricity as power source. These together contribute to the average brightness of the taluk rising to 11 watts/cm^2/μm and average stable lights being 10. Similarly for the taluk of Chandvad in the district of Nashik, the large rural settlement of Darhel contributes to the stable lights and brightness being recorded by the sensor. The taluk of Vaibhawadi in Sindhudurg district has 34 moderately sized villages having 93% of its households with access to electricity. This gives rise to moderate mean brightness and stable lights.

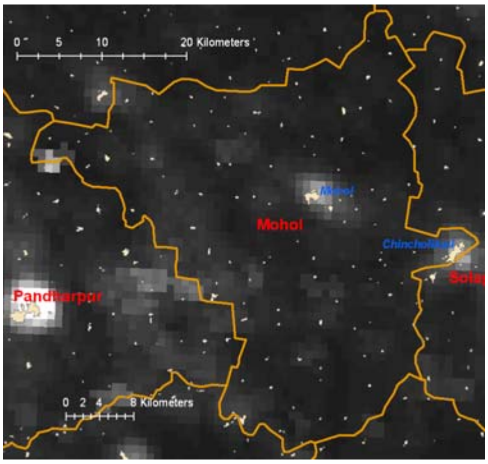


Figure 6B.5 : The taluk of Mohol as seen in DMSP-OLS brightness image. The two main rural settlements are Mohol and Chincholikati. The taluk has many scattered smaller settlements.

6B.2 Development and discussion of the models:

Models were proposed using the mean and standard deviation of brightness and stable lights and the chosen census metrics. Linear regression models and multiple regression models were proposed. The results from the linear regression models are explained in section 6B.2.1 and those from multiple regression models are described in section 6B.2.2.

6B.2.1 Discussion of linear regression models:

Simple linear regression models were calculated using the chosen metrics and mean and standard deviation of stable lights and brightness. Models with and without intercepts were calculated for all the metrics. Adjusted r^2 values were obtained for each model. The equation used in the calculation of adjusted r^2 values is presented in section 5.4.1. The adjusted r^2 values from the linear regression models are presented in table 6B.2.

Table 6B.2 : Adjusted r^2 values from linear regression models. The highest adjusted r^2 values are shaded in light blue and the lowest adjusted r^2 values are shaded in grey

Census Metrics	**Mean brightness with intercept**	**Mean brightness with no intercept**	**SD brightness with intercept**	**SD brightness with no intercept**	**Mean Stable Lights with intercept**	**Mean Stable Lights with no intercept**	**SD Stable Lights with intercept**	**SD Stable Lights with no intercept**
Number of households per square kilometre	0.69	0.75	0.58	0.70	0.48	0.57	0.30	0.52
Total population per square kilometre	0.72	0.78	0.62	0.74	0.51	0.61	0.33	0.56
Female literates per square kilometre	0.70	0.71	0.60	0.70	0.47	0.52	0.31	0.48
Total workers per square kilometre	0.67	0.81	0.56	0.69	0.50	0.69	0.31	0.62
Percentage of households with cars, jeeps and vans	0.15	0.64	0.13	0.39	0.13	0.71	0.08	0.60
Percentage of households with television	0.30	0.74	0.29	0.48	0.35	0.85	0.31	0.77
Percentage of permanent houses	0.15	0.68	0.12	0.39	0.17	0.80	0.13	0.69

Percentage of households using electricity as	0.08	0.66	0.07	0.35	0.07	0.81	0.06	0.69
Urban population per square kilometre	0.73	0.52	0.66	0.65	0.44	0.29	0.31	0.29

The highest and lowest adjusted r^2 as obtained from the models are marked by blue and grey respectively in table 6B.2. It was observed that model without intercept using mean stable lights and brightness outperformed all the other models. These models produced the highest adjusted r^2 values for all the metrics. On the other hand, the model with intercept including the standard deviation of stable lights exhibited the lowest r^2 values. From table 6B.2, it was noted that models with no intercepts exhibited higher r^2 values compared to those with intercepts for all the metrics except *urban population per square kilometre*. Although some of the adjusted r^2 values were very low (for example 0.12 for percentage of permanent census houses in the model with standard deviation of brightness with intercept), all were significant at the 95% confidence interval. For models using mean and standard deviation of brightness and mean stable lights with intercept, the adjusted r^2 values were higher for the demographic metrics than for the amenities datasets. For demographic metrics these values ranged from 0.29 for *urban population per square kilometre* to 0.81 for number of *total workers per square kilometre*. However, for the amenities datasets, the value ranged from 0.06 for *percentage of households with access to electricity* as power source to 0.85 for *percentage of households having television*. Although wide range of variation was noted in the r^2 values for different models, those using mean brightness with no intercepts exhibited higher r^2 values for demographic metrics while model with mean stable lights without intercept produced higher adjusted r^2 for amenities datasets. On the other hand, the model using standard deviation of stable lights with no intercept had the lowest r^2 values for most of the metrics.

6B.2.2 Discussion of multiple regression models:

The multiple regression models were calculated with more than one predictor variables. Models were tested with a) Mean and standard deviation stable lights; b) mean and standard deviation of brightness and c) mean and standard deviation of both stable lights and brightness together. Similar to linear regression models, these models were also calculated both with and without intercepts. The method of backward elimination was used to calculate the multiple regression models. Details of the method of backward elimination are explained in section 5.4.2.

Some of the variables were pooled out from the multiple regression models by the method of backward elimination. The variables that were pooled and included in the multiple regression models are listed in table 6B.3. Standard deviation of brightness was pooled in the model A. For all the other models (B – F), the independent variables were included in the calculation.

Table 6B.3: Variables pooled and included in multiple regression models

Independent variables in the models	**Variables pooled in stepwise regression**	**Variables included in multiple regression**
A) Mean and SD brightness and stable lights with intercept	standard deviation brightness;	mean brightness; mean stable lights; standard deviation stable lights
B) Mean and SD brightness and stable lights with no intercept	none	mean brightness; mean stable lights; standard deviation stable lights; standard deviation brightness
C) Mean and SD of brightness with intercept	none	mean brightness; standard deviation brightness
D) Mean and SD of brightness with no intercept	none	mean brightness; standard deviation brightness
E) Mean and SD stable lights with intercept	none	mean stable lights; standard deviation stable lights
F) Mean and SD stable lights with no intercept	none	mean stable lights; standard deviation stable lights

The highest and the lowest values are marked in blue and grey respectively. The adjusted r^2 values were all significant at the 95% confidence interval. The values were observed to be higher than those obtained from linear regression models (table 6B.2). The model with mean and standard deviation of brightness and stable lights with no intercept outperformed all the other models. The model demonstrated the highest adjusted r^2 for all the census metrics. The values ranged from 0.73 for *percentage of households with cars, jeeps and vans* to 0.89 for *percentage of households with access to electricity*. The model with intercept using mean and standard deviation of stable lights had the lowest adjusted r2 values for all the metrics.

Table 6B.4 : Adjusted r^2 values from multiple regression models. The highest adjusted r^2 values are shaded in light blue and the lowest adjusted r^2 values are shaded in grey

Census Metrics	**Mean and SD brightness and stable lights with intercept**	**Mean and SD brightness and stable lights**	**Mean and SD of brightness with intercept**	**Mean and SD of brightness**	**Mean and SD stable lights with intercept**	**Mean and SD stable lights**
Number of households per square kilometre	0.72	0.80	0.72	0.75	0.53	0.57
Total population per square kilometre	0.75	0.83	0.74	0.79	0.55	0.61
Female literates per square kilometre	0.73	0.79	0.72	0.73	0.52	0.52
Total workers per square kilometre	0.71	0.83	0.71	0.81	0.58	0.70
Percentage of households with cars, jeeps and vans	0.15	0.73	0.16	0.73	0.14	0.73
Percentage of households with television	0.35	0.86	0.30	0.83	0.35	0.86
Percentage of permanent houses	0.17	0.84	0.16	0.82	0.17	0.82
Percentage of households using electricity as power source	0.07	0.89	0.07	0.87	0.06	0.85
Urban population per square kilometre	0.78	0.78	0.73	0.65	0.46	0.29

Both the linear and the multiple regression models were validated over the withheld taluks. The results from the validation are described in the following section.

6B.3: Model Validation:

The models were validated using the withheld 90 taluks. The mean and standard deviation of stable lights and brightness from these withheld taluks were used to validate the linear regression and multiple regression models. Section 5.5 describes the formula used to calculate the difference in the predicted values between the models and the census values.

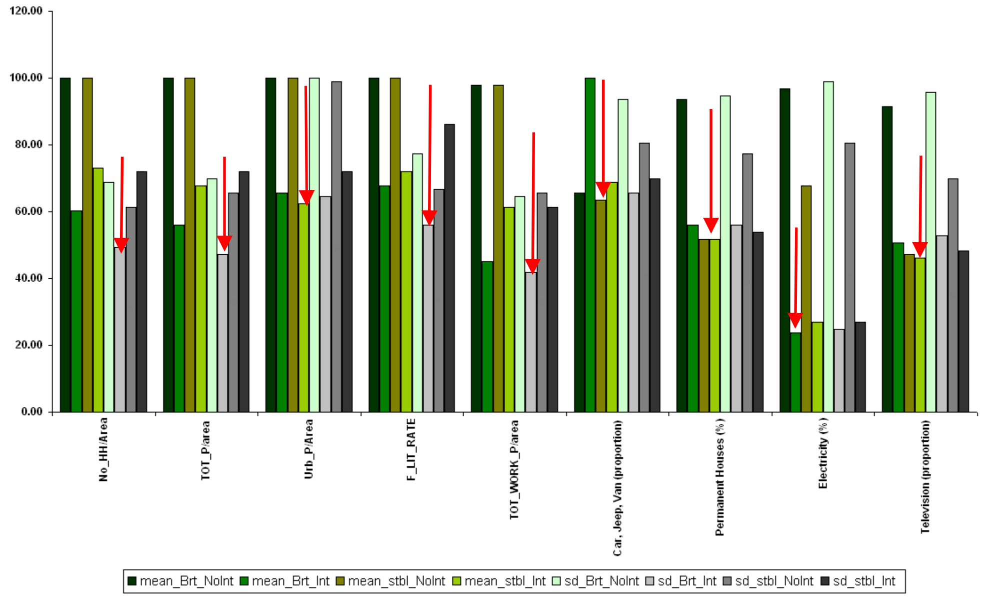


Figure 6B.6 : Percentage of taluks with predicted values beyond |25|% error margin for each linear regression model. The red arrows indicate the best performing models.

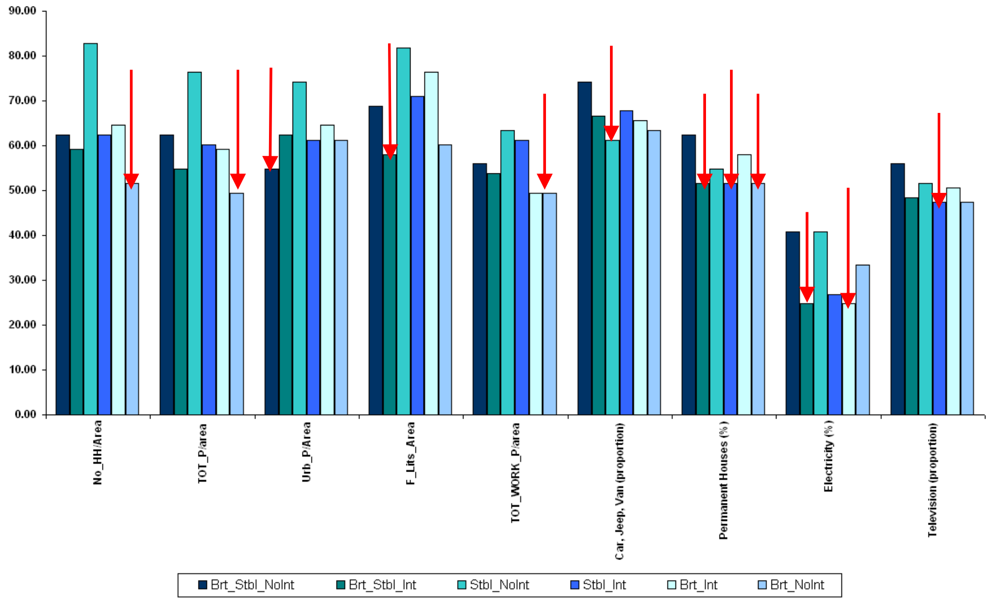
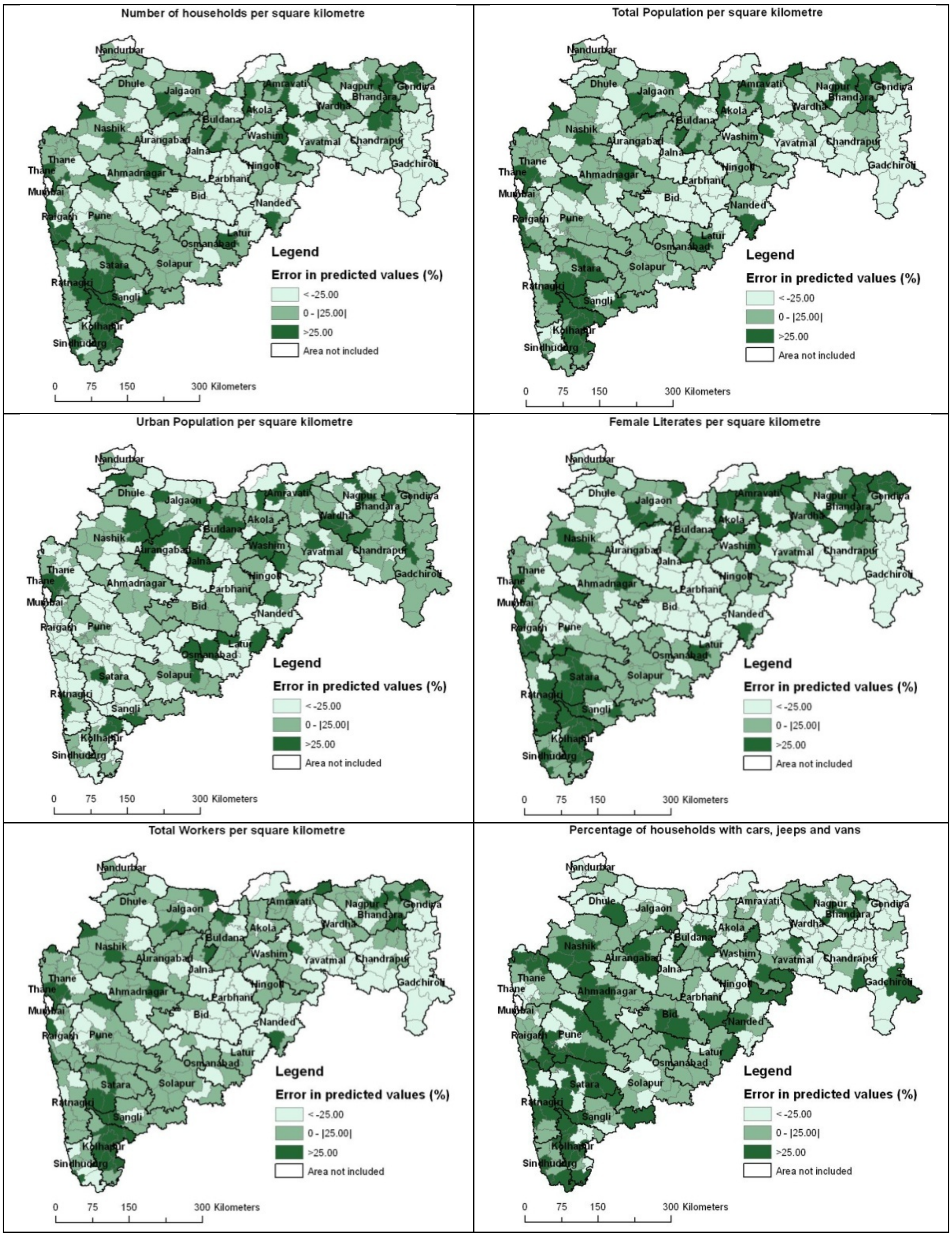


Figure 6B.7 : Percentage of districts with predicted values beyond |25|% error margin for each multiple regression model. The red arrows indicate the best performing models.

The models which most accurately predicted the values within a |25|% difference of error were selected. Figure 6B.6 demonstrates the results from validation of the linear regression models and figure 6B.7 illustrates the results from the multiple regression models.

Success in the predicted census variables using the models varied. In the best case scenario, 60% of the Taluks were successfully predicted within the chosen |25|% error margin. Even for the poorest performing models, 20% of the Taluks were predicted within the error margin. The linear regression model with intercept using standard deviation of brightness as the predictor variable outperformed the other models. This model optimally predicted four out of nine metrics. These include *number of households per square kilometre, total population per square kilometre, number of female literates per square kilometre* and *number of workers per square kilometre*. Unlike the models at the district level, most of the census parameters could be predicted with one optimum model at the taluk level. However, for some of the metrics there was more than one optimal model. A variable such as percentage of permanent census houses was predicted within the 25% error margin by four optimum models, marked by red arrows in figure 6B.6. In case of such metrics, the adjusted r^2 value was used to determine the most suitable model.

The chosen optimum models were used to predict the census metrics. The models are shown in table 6B.5. Residual values were calculated and mapped to show the percentage of difference between the predicted values and the census recorded metrics. The residual maps are shown in fig 6B.8 and the maps with the predicted metrics are shown in fig 6B.9. A residual value is defined as the difference between the recorded value and the value predicted by the models. It is sometimes referred to as the unexplained difference between the two values. In this study the residual values are expressed as percentage and are calculated using the formula explained in section 5.5.



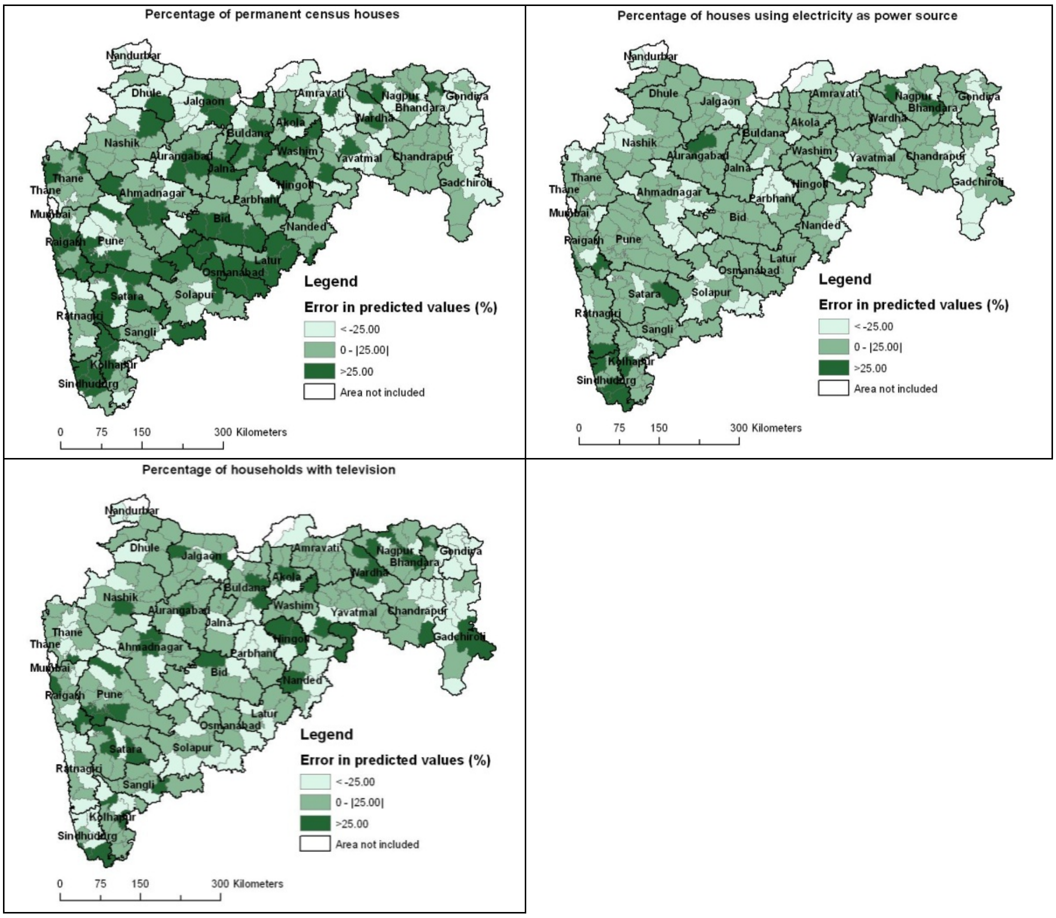


Figure 6B.8 : Maps showing percentages of error in the predicted values from the models for the taluks

The maps showing the residual values of the census metrics are shown in figure 6B.8. The taluks which were under-predicted by the models had more than 25% error and are shown in dark shades. The taluks for which the models predicted higher than the census recorded data, gave rise to negative errors and are shown by light shades. The models predicted all the metrics within the |25|% error margin for more than 50% of the taluks. The *percentage of households with access to electricity* was predicted optimally for more than 70% of the taluks. There were patches of over and under predictions all over the state. This shows that there are regional variations in the levels of development all over Maharashtra.

The demographic metrics such as *number of households per square kilometre*, *total population per square kilometre* and *total workers per square kilometre* displayed similar trends of error distribution. These metrics were under-predicted for the taluks in the south eastern part of the state in the districts of Kolhapur, Sangli and Ratnagiri. The errors were more than 25% in these areas. However, for majority of the taluks in the districts of Gadchiroli, Parbhani and Jalna, the demographic metrics were over-predicted by the models. These taluks displayed error of less than 25%. Number of *female literates per square kilometre* was over-predicted for most taluks of Aurangabad, Parbhani, Bid, Jalna, Chandrapur and Gadchiroli. Taluks in districts such as Kolhapur, Satara, Ratnagiri, Amravati, Akola and Gondiya exhibited more than 25% difference between the census recorded and the predicted values using the model. The metrics on amenities were also over and under-predicted for some of the taluks all over the state. *Percentage of households with access to electricity* was predicted accurately for most of the taluks. It was over-predicted for some of the western taluks in Gadchiroli and under-predicted for few taluks in Sindhudurg, Kolhapur and Ratnagiri in the southern part of Maharashtra. Variable such as *percentage of households with cars, jeeps and vans* was under-predicted for most of the taluks in western and southern districts such as Nashik, Ahmadnagar, Satara, Ratnagiri and Kolhapur. Percentage of permanent census houses was under-predicted with error of more than 25% for some taluks in the eastern part of the state in the districts of Bid, Osmanabad and Latur.

6B.4: Results and Discussion:

6B.4.1: Selected Models and the mapping of census metrics:

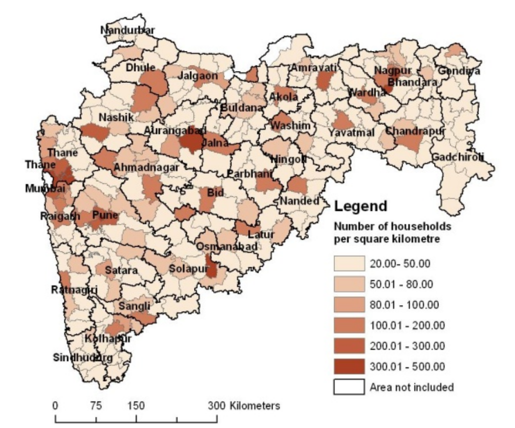
From the results of the residuals maps, the optimum models were shortlisted. The chosen models are shown in table 6B.5.

Table 6B.5 : Optimum Models

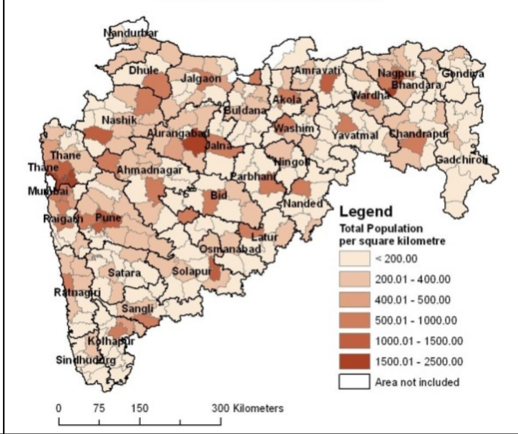
Census Metrics	Chosen Models
Number of households per square kilometre	16.34 + 5.11 x ‘standard deviation brightness’
Total population per square kilometre	86.64 + 24.17 x ‘standard deviation brightness’
Urban population per square kilometre	15.62 x ‘mean brightness’ +12.22 x ‘standard deviation brightness’– 1.36 x ‘mean stable lights’ – 26.79 x ‘SD stable lights’
Female literates per square kilometre	12.96 + 8.21 x ‘standard deviation brightness’
Total workers per square kilometre	59.48 + 7.92 x ‘standard deviation brightness’
Percentage of households with cars, jeeps and vans	0.28 x ‘mean stable lights’ – 0.19 x ‘standard deviation stable lights’
Percentage of permanent houses	5.33 x ‘mean brightness’ – 2.95 x ‘SD brightness’
Percentage of households using electricity as power source	63.35 + 0.51 x ‘mean brightness’
Percentage of households with television	12.24 + 1.69 x ‘mean stable lights’

The maps showing the predicted values obtained from the optimum models (table 6B.5) are shown in figure 6B.9. The maps show the variation in the distribution of census metrics between the urban, sub-urban and rural parts of the districts.

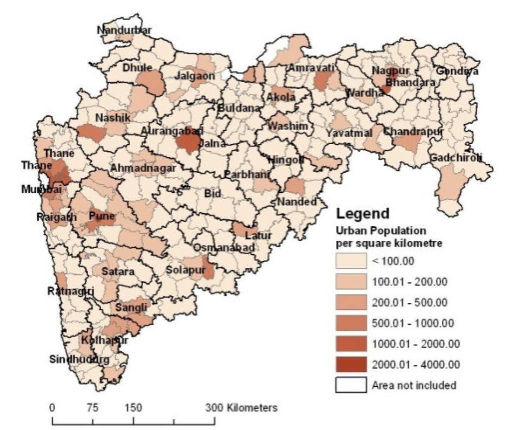
Number of households per square kilometre



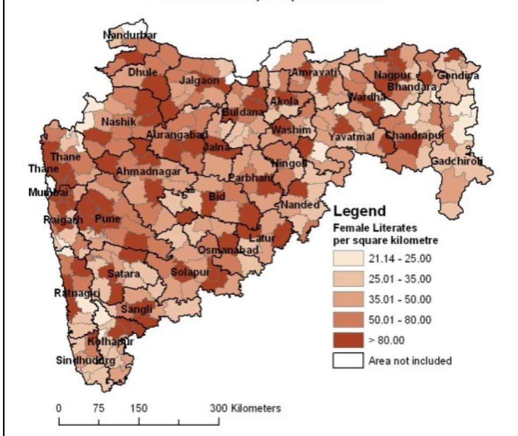
Total Population per square kilometre



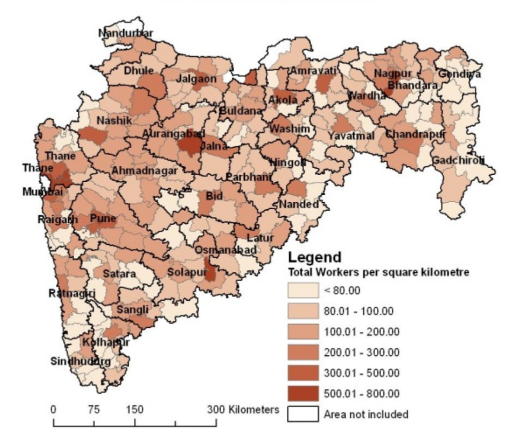
Urban Population per square kilometre



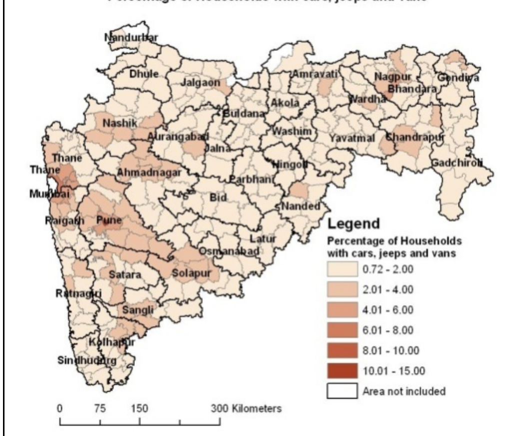
Female Literates per square kilometre



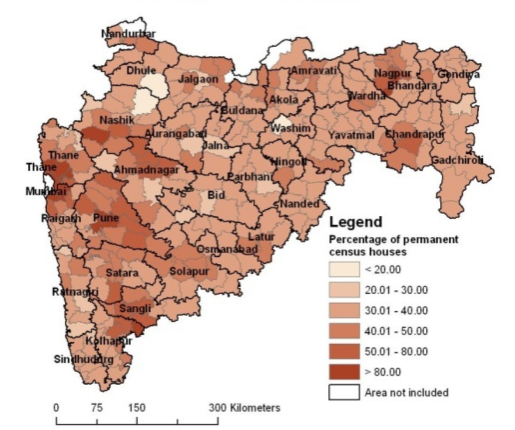
Total Workers per square kilometre



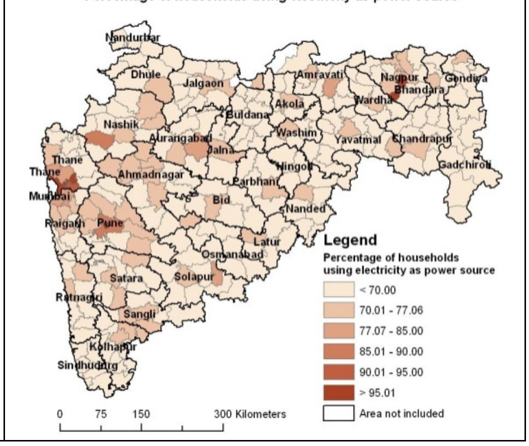
Percentage of Households with cars, jeeps and vans



Percentage of Permanent Census Houses



Percentage of households using electricity as power source



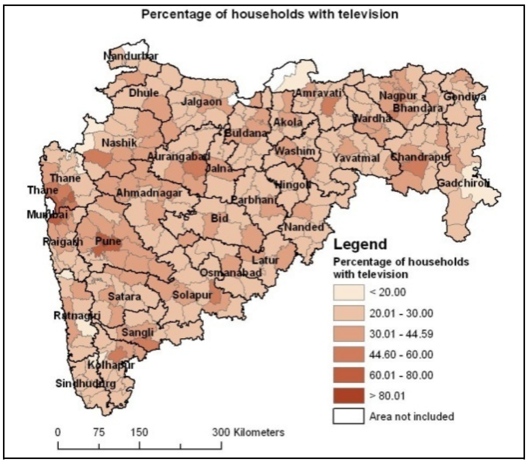


Figure 6B.9 : Predicted maps of census metrics for taluks of the state of Maharashtra using the proposed models

The maps in figure 6B.9 show the following patterns:

- Every district consisted of taluks with both high and low values for most of the metrics.
- There were taluks with both high and low *number of households per square kilometre* in all the districts. However, Gadchiroli displayed a uniform low distribution with 20 – 50 households per square kilometre. This made it one of the most rural districts of Maharashtra. Similar trend was also noted for *total population per square kilometre*. Gadchiroli had low *population density* with less than 200 persons per square kilometre for most of the district. High and moderate *population density* was noted in the western part of Maharashtra in the districts of Pune, Ratnagiri, Raigarh, Thane and Ahmadnagar and in Nagpur in northern part of the state.
- *Urban population per square kilometre* was low (less than 100 *urban population per square kilometre*) for most of the state except for taluks near cities such as Pune, Aurangabad, Thane and Nagpur.
- Taluks with more than 80 *female literates per square kilometre* were found in all the districts. This indicated ongoing human development all over the state of Maharashtra. Districts such as Hingoli and Gadchiroli, however, demonstrated moderate distribution of *female literates per square kilometre*.
- High values of *total workers per square kilometre* (more than 800) were found in taluks near big urban centres of the state such as Nagpur, Aurangabad, Solapur, Pune and Thane. However, there was a moderate distribution of *total workers per square kilometre* in almost all the districts (values ranging from 100 – 300 workers per square kilometre). The map showed levels of employment in all the sectors all over Maharashtra.

- Less than 2% of the households had access to cars, jeeps and vans all over the state. Around 10 – 15% of the households had cars, jeeps and vans in the taluks near large urban areas such Pune, Ahmadnagar, Thane, Raigarh and Solapur.

- Most of the districts had more than 40% of permanent census houses, with electricity access to around 70% of the households for most of the state. Urban taluks such as Thane, Pune and Nagpur displayed more than 95% of the households using electricity as power source. The spread of electrification in both rural and urban areas of Maharashtra supported around 30 – 45% of the households in most taluks to use television.

The taluks are subdivided into villages. These are the smallest administrative units in the country. Information obtained from global composite DMSP-OLS images were also used to propose models for villages. The correlations and models at the village level are described in the following part (6C).

6C. Surrogate Census at Village Level:

Villages are the smallest administrative units in India. These are rural units of the country. Like any other Indian state, the taluks in the state of Maharashtra are divided into smaller administrative units called villages. The census accounted for more than 49000 villages for the state of Maharashtra while the data from SAC provided information on 43865 village units. Of these, 65 units were described as either a creek or a forest. As a result, this study was restricted to the available 43800 villages. Details of the data quality issues encountered in this study are explained in section 4.7.

6C.1 Sampling and pre-processing of datasets:

Since the main objective of this study at the village level was to assess the applicability of DMSP-OLS images at a small spatial scale in predicting census metrics, a stratified random sample of 100 villages was selected from the available 43800. Details of the method of stratified random sampling are explained in section 4.4. The method of stratification and village selection were carried out on the basis of the following steps:

1. The districts were sorted on the basis of their total population. The district of Mumbai (suburban) had the highest number of people followed by the districts of Thane, Pune and Nashik respectively. The district of Nagpur had the fifth largest population in the state. On the other hand, the district of Sindhudurg was the least populated district of the state followed by the districts of Gadchiroli, Hingoli and Washim. The districts of Amravati, Sangli, Yavatmal, Buldana and Raigarh had moderate total population. Three districts were selected representing the range of total population distribution. The three chosen districts were Pune, Yavatmal and Gadchiroli. Pune had a total

population of more than 7 million, followed by Yavatmal with a population of 2.5 million and Gadchiroli with a population of less than 1 million was the least populated.

2. In total 4688 villages were present in the districts of Pune, Yavatmal and Gadchiroli. Of these altogether 90 villages were chosen randomly (30 from each district). From these 90 villages, 60 units were randomly sampled (20 from each district) to propose the models. 10 villages from each district were withheld for validation.

3. Mean and standard deviation of brightness and stable lights were calculated for each village.

4. In order to compare the utility of DMSP-OLS images at village level with larger spatial units (districts and taluks), attempt was made to choose the same demographic and amenities datasets as districts and taluks. However, data on only three demographic variables were available at the village level from the census. These were: *number of households*, *total population* and *total number of workers*. The densities (per square kilometre) of these variables were calculated for the analyses.

6C.2 Discussion of correlations:

Correlations between the three census metrics and mean and standard deviation of brightness and stable lights were calculated for the 60 villages. The metrics displayed very low correlations (less than 0.3) with stable lights and brightness at the 95% confidence interval. At this confidence interval, there were significant correlations only between census metrics and mean stable lights. An attempt was made to study the correlations at the 90% confidence interval. At the confidence interval of 90%, significant correlations were found to exist between census metrics and means of stable lights brightness. Villages being small in areas (ranging from a few square kilometres to several square kilometre) there was absence of high deviations in brightness and stable lights within a village. As a result, there were no significant correlations between standard deviation of brightness and stable lights and the census metrics at the village level. The correlations at the 90% confidence interval are shown in figure 6C.1.

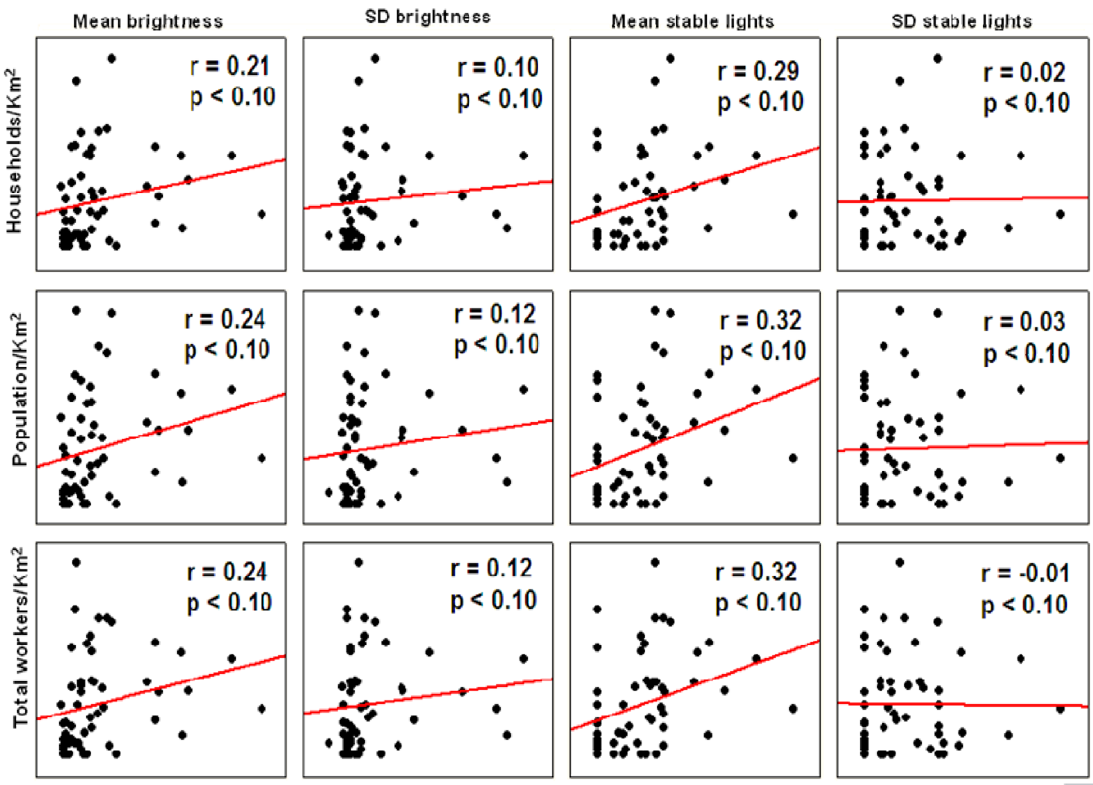


Figure 6C.1: Correlations between census metrics and mean and standard deviation of brightness and stable lights for villages as recorded from global composite image of 2001

6C.3 Models:

From the results of the correlations, models for surrogate census at the village level were proposed using the mean of brightness and stable lights at the 90% confidence interval. Both linear regression models and multiple regression models were tested. The results from these models are explained in the following sections.

6C.3.1 Discussion of Linear regression models:

The adjusted r^2 values as obtained from linear regression models with intercepts were not significant for the villages. Therefore, only the models without intercepts were considered. The adjusted r^2 values from these models are presented in table 6C.1.

Table 6C.1 : Adjusted r^2 values from linear regression models

Census Metrics	Mean brightness with no intercept	Mean Stable Lights with no intercept
Number of households per square kilometre	0.51	0.44
Total population per square kilometre	0.54	0.47
Total workers per square kilometre	0.54	0.47

It was found that the correlations between the census metrics and mean brightness were higher compared to those with mean stable lights. These were calculated at the 90% confidence interval.

6C.3.2 Discussion of multiple regression models:

Multiple regression models with and without intercepts were calculated. Means of stable lights and brightness were used as predictor variables. Models were tested both with and without intercepts. The variables which were pooled and included in the multiple regression models are shown in table 6C.2.

Table 6C.2 : Variables pooled and included in multiple regression models

Independent variables in the models	Variables pooled in stepwise regression	Variables included in multiple regression
A) Means of brightness and stable lights with intercept	Mean brightness	Mean stable lights
B) Means of brightness and stable lights with no intercept	Mean stable lights	Mean brightness

Mean brightness was pooled in the model with intercept in the process of backward elimination. On the other hand, mean stable light was removed from the model with no intercept. As a result only one variable was used in the models. Therefore, there was no multiple regression model at the village level.

6C.4 Model validation:

In the absence of multiple regression models, the linear regression model with mean brightness with no intercept was chosen as optimum models. These models demonstrated higher adjusted r^2 values compared to those with mean stable lights.

The chosen models were used to predict the census metrics for the villages. The chosen models for the census metrics at the village level are shown in table 6C.3. Residual values were calculated and mapped to show the percentage of difference between the predicted values and the census recorded values.

The maps showing the residual values of the census metrics for the villages of Pune, Yavatmal and Gadchiroli are shown from figure 6C.2 to 6C.4. The villages which were under-predicted by the models exhibited more than 25% error and are shown in dark shades. On the other hands, the villages for which the models predicted higher than the census recorded data, showed negative errors and are denoted by light shades in figure 6C.2 to 6C.4. The models predicted all the metrics with moderate accuracy, with around 27% to 38% of the villages being predicted within the |25|% error margin.

In the district of Pune, the *number of households per square kilometre* was predicted within the error margin for 27.6% of the villages (figure 6C.2). These villages were located in patches mainly in the north- western and southern part of the district. The metrics were over-predicted for the villages on the western part of the district and the central part around the city of Pune. These villages showed negative errors as the predicted values were higher than those recorded by the census. Patches of over-prediction were also noted for villages along the north-eastern and eastern part of the district. Error of more than 25% was found in the villages mainly in the southern part of Pune. *Total population per square kilometre* was optimally predicted for 26.8% of the villages in Pune. *Total workers per square kilometre* was predicted optimally for 26.9% of the villages. The distribution of the villages for which these two metrics were over and under-predicted showed the same trend as *number of households per square kilometre*. These metrics were also over-predicted for the villages in the western part of the district and areas around Pune. The pattern of error distribution showed that parts of the districts near the urban centres which recorded higher mean brightness and stable light values tended to be over-predicted by mostly all models.

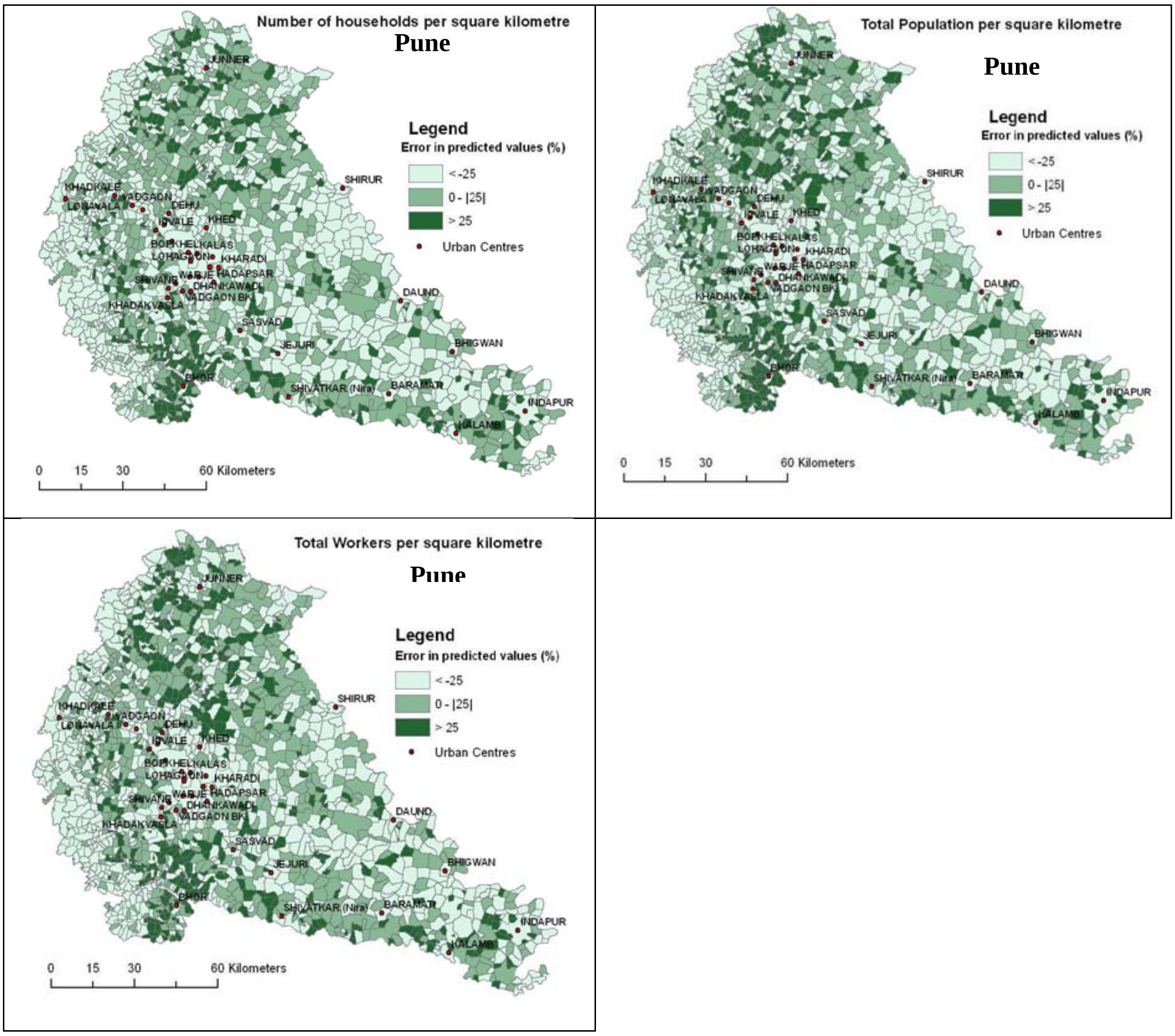


Figure 6C.2 : Maps showing percentages of error in the predicted values from the models for the villages of Pune

For the district of Yavatmal, *number of households per square kilometre*, *total population per square kilometre* and *total workers per square kilometre* were predicted optimally for 37% to 38% of the villages. These villages were distributed all over the district with no distinct pattern of distribution.

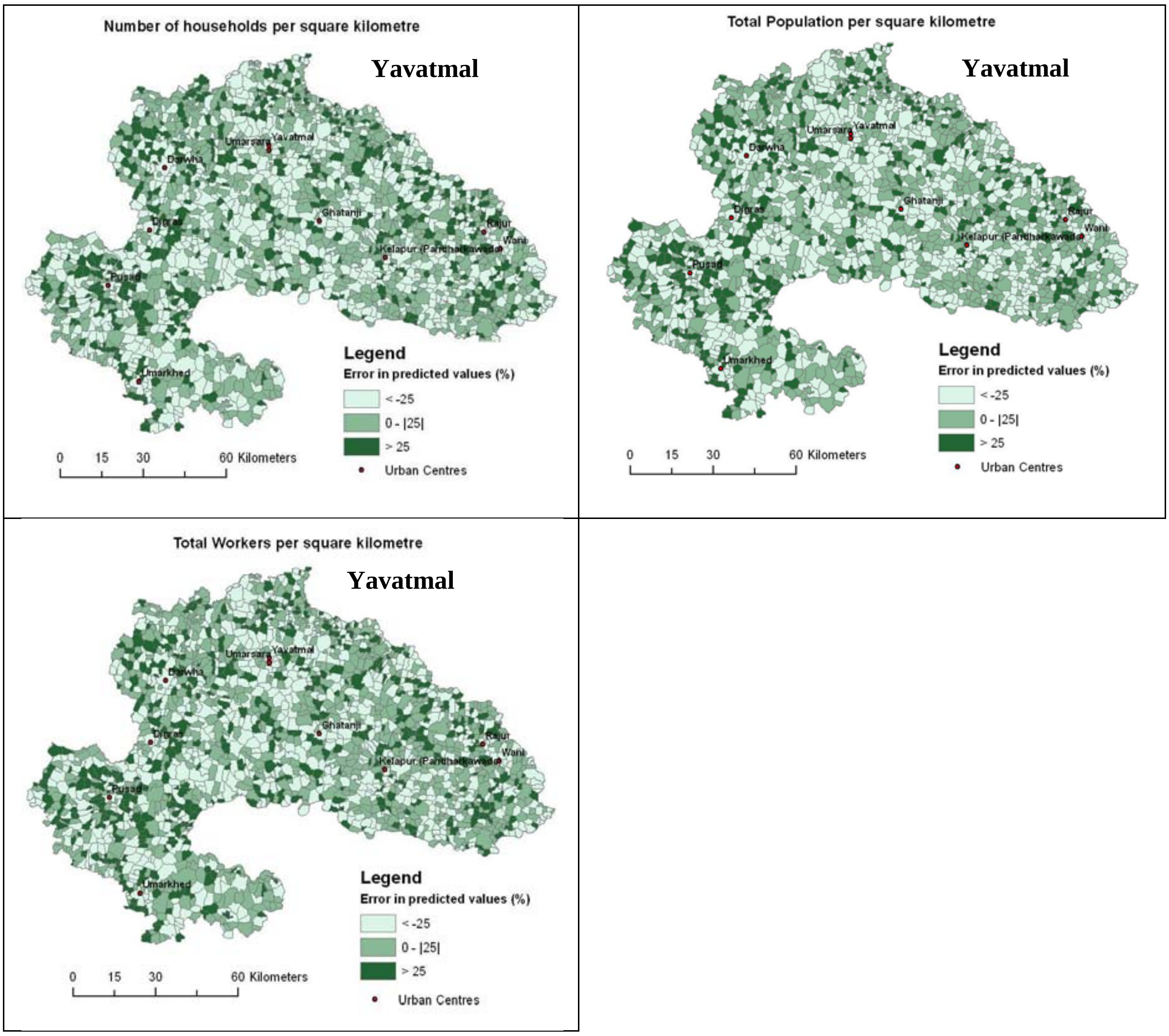


Figure 6C.3 : Maps showing percentages of error in the predicted values from the models for the villages of Yavatmal

For the district of Gadchiroli, all the metrics were predicted optimally for approximately 27% of the villages. The western part of the district is under forest and recorded no brightness or stable lights information. However, the urban centres of Gadchiroli, Aheri and Desaiganj are located in this part of the district. The metrics were predicted within the error range of |25|% for the villages around these urban areas. The metrics were over-predicted for the villages in the eastern part of Gadchiroli. These villages recorded negative errors and are shaded in light green in figure 6C.4. The villages for which the census metrics were under-predicted recorded positive errors of more than 25%. These villages were distributed in patches mainly in the western and northern part of Gadchiroli.

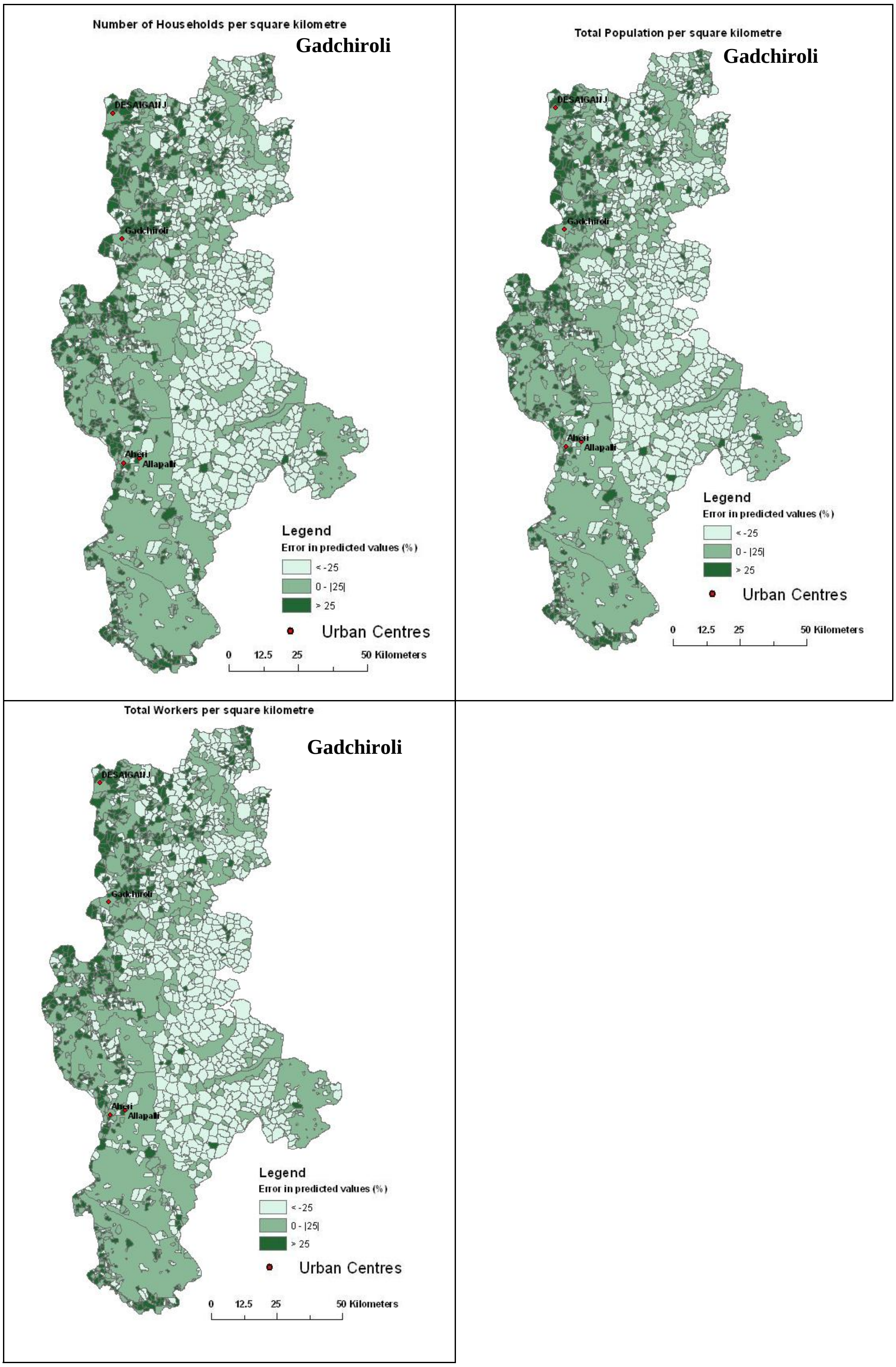


Figure 6C.4 : Maps showing percentages of error in the predicted values from the models for the villages of Gadchiroli

6C.5 Chosen models and predicted maps:

The models used for prediction of metrics for the villages are shown in table 6C.3.

Table 6C.3: Optimum models

Census Metrics	Chosen Metrics
Number of households per square kilometre	3.05 x 'mean brightness'
Total population per square kilometre	15.54 x 'mean brightness'
Total workers per square kilometre	7.18 x 'mean brightness'

The maps showing the predicted values obtained from the optimum models (table 6C.3) are shown in figures 6C.5 to 6C.7. The maps reveal the following patterns of distribution of the census metrics across the villages in the districts of Pune, Yavatmal and Gadchiroli:

- In Pune higher densities of number of households, total population and total workers are found in areas around the urban centres of Vadgaon Bk, Hadapsar, Khed, Kharadi, Kivale and Dehu in the central part of the district. Higher densities are also noted around the urban centres of Jumner, Shirur, Baramati, Kalamb and Bhor. Areas with a moderate *number of households per square kilometre* and *total population per square kilometre* extend linearly northward and eastward from the cluster of cities in the centre of the district. However, *total workers per square kilometre* are distributed evenly in the district towards the east and northern part. The maps show that although population and households concentrate more towards cities and towns, level of employment is prevalent all over the district.
- In the district of Yavatmal, villages with high densities of census metrics are found around the urban centres of Yavatmal, Umarsara, Darwa, Digras, Pusad, Kelapur, Ghatanji and Umarkhed. Most of the villages in the district have around 20 - 40 *number of households per square kilometre* with around 100 – 200 people per square kilometre. Higher densities of *total workers per square kilometre* (50 – 100) are found in clusters in the southern and eastern part of Yavatmal. The rest of the district has around 30 – 50 workers per square kilometre.
- Compared to the other two districts, Gadchiroli is mainly a rural district with only three towns of Desaiganj, Gadchiroli and Aheri. However, contrary to Pune and Yavatmal, the areas around these towns have very low densities of all the three census metrics. No villages in the district exhibit high values for the metrics. *Number of households per square kilometre* ranges from 1 – 30 for majority of the villages. The *population density* varies from 1 – 150 persons per square kilometre for the villages. However, there are some patches of villages with comparatively high *total workers per square kilometre*. These villages have around 50 – 100 workers per square kilometre and are located in isolated patches in the eastern and northern

part of Gadchiroli. Most of the villages in the district have around 30 – 50 workers per square kilometre. There are large tracts of villages with no population and household density in the western part of Gadchiroli.

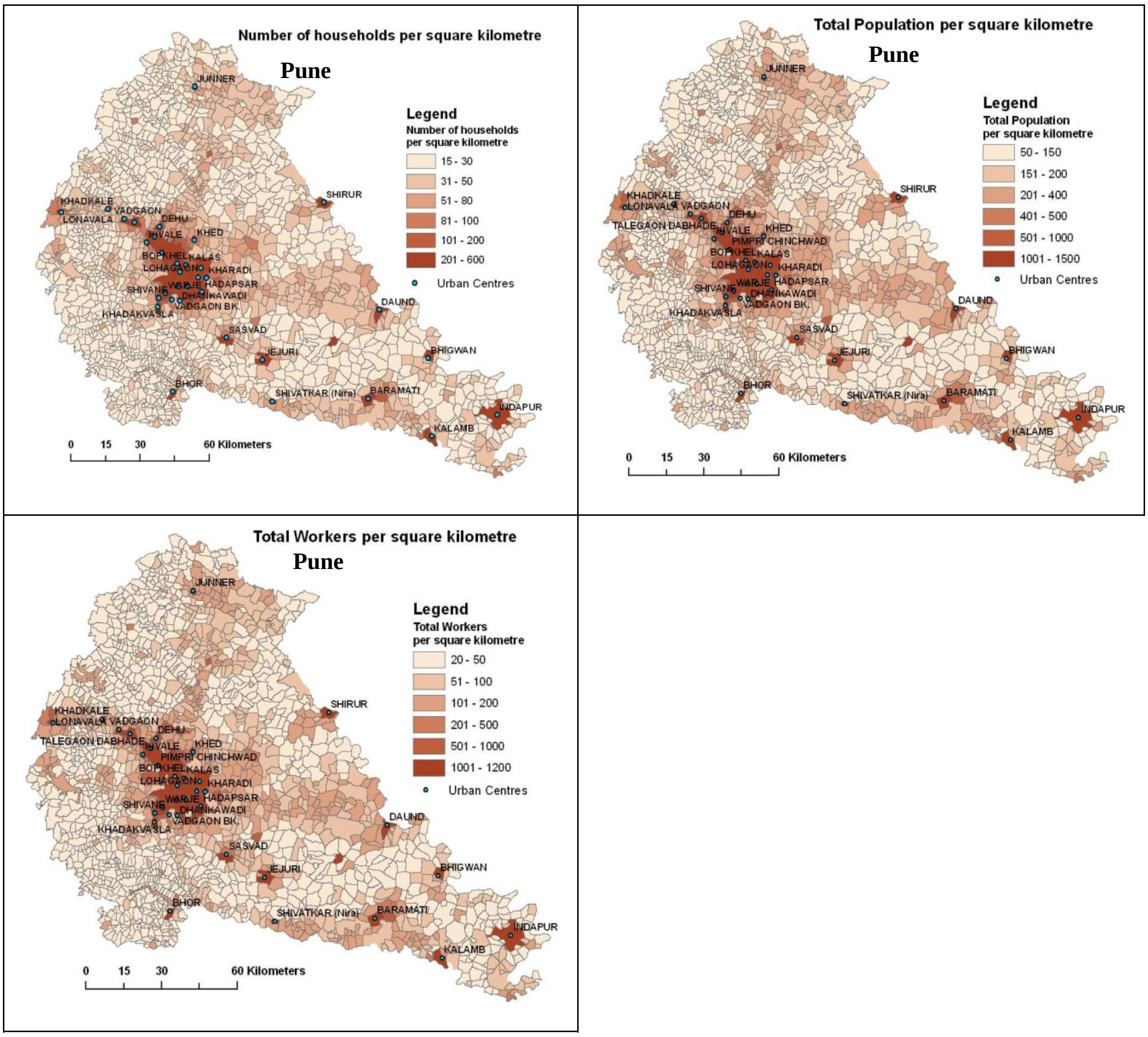


Figure 6C.5 : Predicted maps of census metrics for villages of Pune using the proposed models

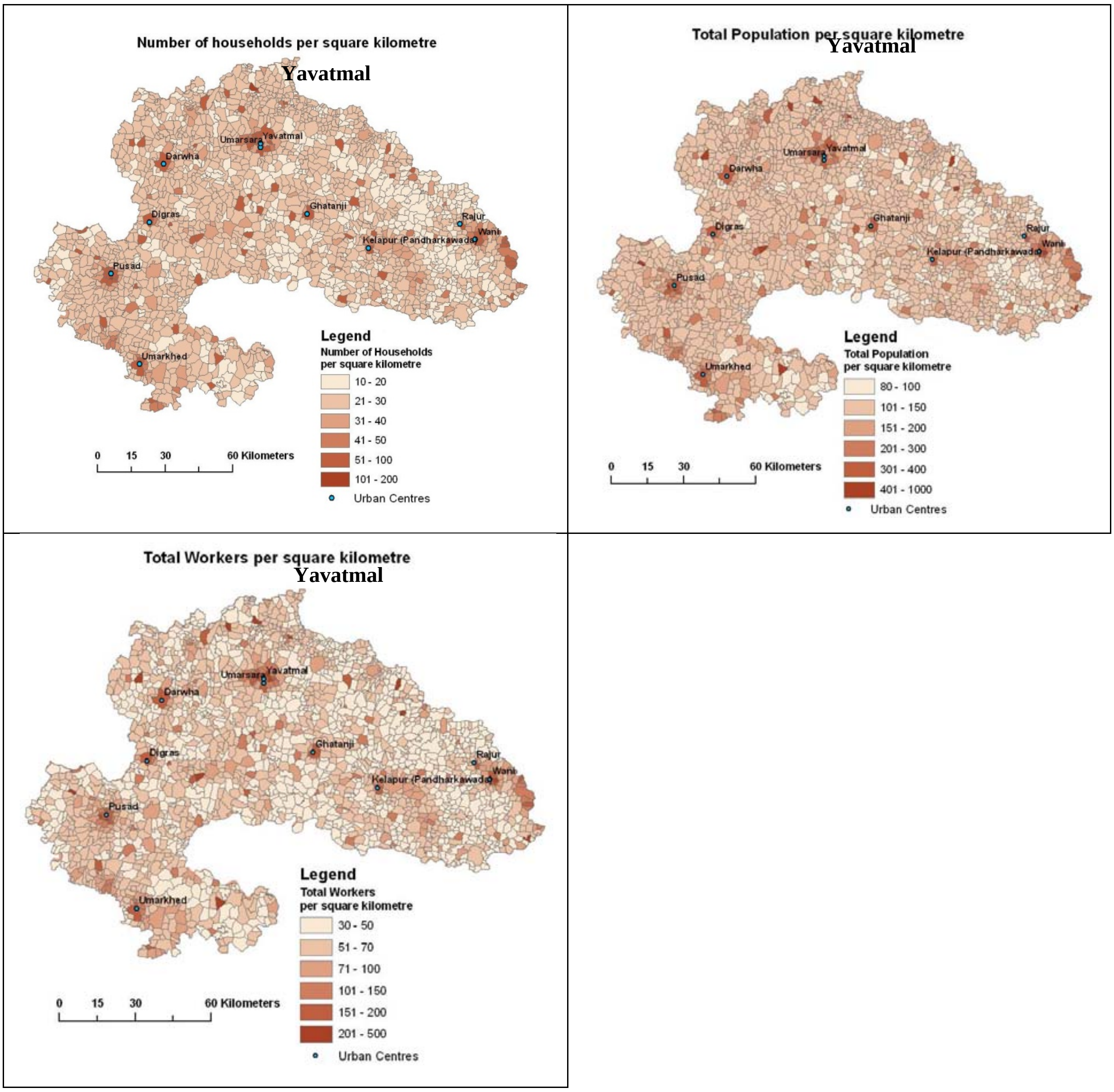


Figure 6C.6 : Predicted maps of census metrics for villages of Yavatmal using the proposed models

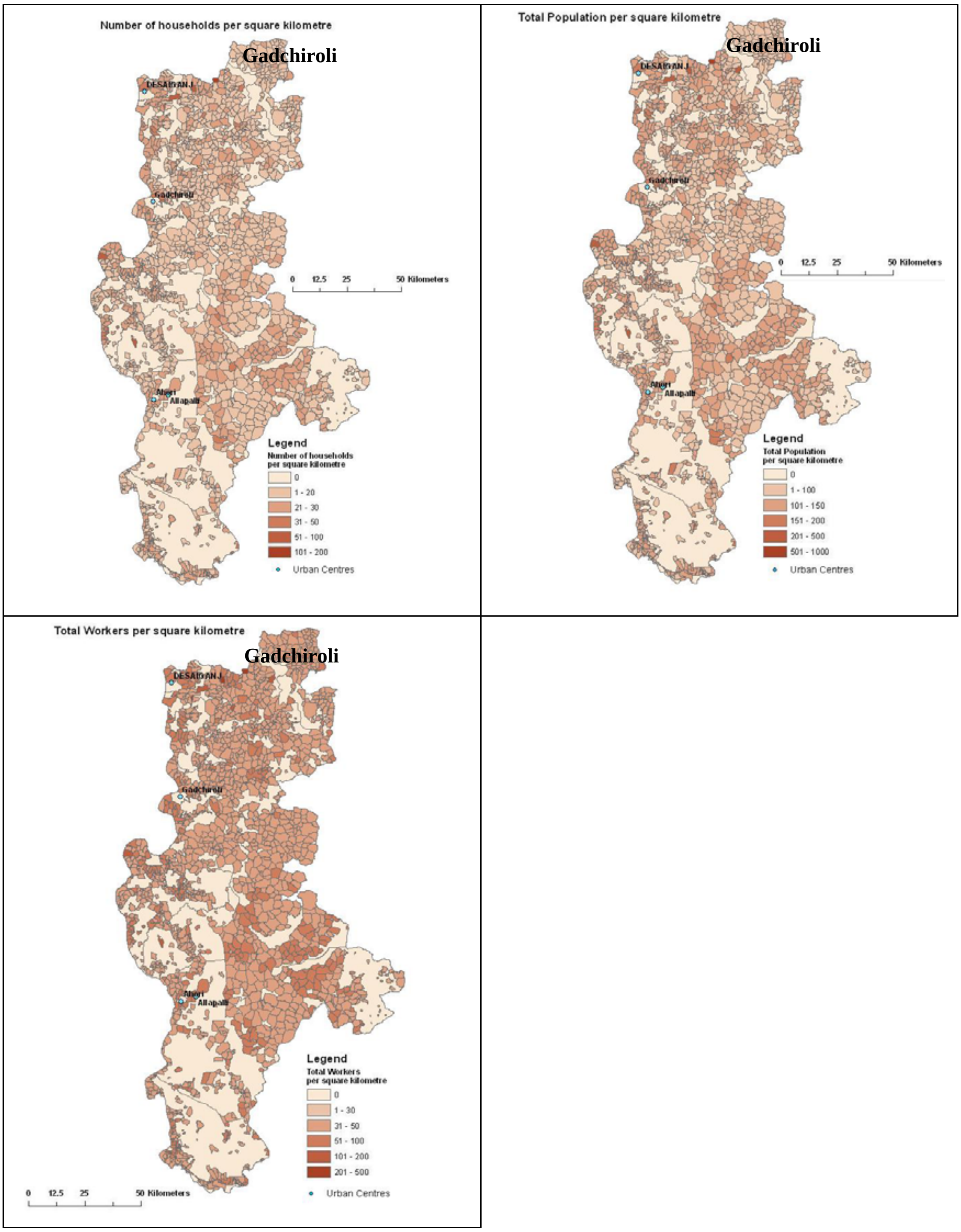


Figure 6C.7 : Predicted maps of census metrics for villages of Gadchiroli using the proposed models

6.1: Summary:

The stable lights and the brightness datasets from DMSP-OLS images were used to propose the models at district, taluks and villages for the state of Maharashtra. The summary of the results obtained from these models are presented in this section.

Table 6.1 : Summary of results at districts, taluks and villages

DISTRICTS	TALUKS	VILLAGES
Ten census metrics were used in the analyses. Correlation coefficients ranged from 0.3 to 0.8 with stable lights and 0.4 -0.9 with brightness at the 95% confidence interval. Overall, correlations of the metrics with brightness were greater than those obtained with stable lights. The linear regression model with mean brightness with no intercept predicted the best results. The multiple regression model with mean and standard deviation of stable lights and brightness with no intercept outperformed the others. Overall the multiple regression models performed better than linear regression models. The models optimally proposed metrics for 70% of the districts.	Nine census metrics were used in the analyses. There was no economic indicator available from the census Correlation coefficients ranged from 0.4 to 0.7 with stable lights and 0.3 -0.8 with brightness at the 95% confidence interval. Overall, mixed results were observed in the correlations of the metrics with brightness and stable lights. The linear regression models with mean brightness and mean stable lights with no intercept predicted the best results. The multiple regression model with mean and standard deviation of stable lights with no intercept outperformed the others. Overall the multiple regression models performed better than linear regression models. The models optimally proposed metrics for 20% - 70% of the taluks.	Three census metrics were used in the analyses. Remaining indicators were not available from the census. Correlations between census metrics and DMSP-OLS image derived information were not significant at the 95% confidence interval. So these were calculated at the 90% confidence interval. Correlation coefficients were lea than 0.3 at the 90% confidence interval. Correlations were significant only with the means of brightness and stable lights. The linear regression models with mean brightness with no intercept predicted the best results. There was no multiple regression model for the villages. The models optimally proposed metrics for 25% - 40% of the villages.

On comparing the results at all the three spatial scales, it was observed that for districts, the correlation coefficients (r) ranged from 0.3 to 0.9 at the 95% confidence interval. The correlations were, however, not significant for percentage of households with access to electricity and stable lights. At the taluk level, the correlations ranged from 0.3 to 0.8 at the 95% confidence interval. The correlations were significant for all the census metrics at this spatial scale. For villages, there were significant correlations between census metrics and means of brightness and stable lights only at the 90% confidence interval. The values of the correlation coefficients were below 0.3. Overall, the correlations were higher at the districts than at the taluks. Therefore it was observed that the best correlations were obtained for the districts and then the taluks. As a result, it answers research question 2 (section 1.2) that the DMSP-OLS images are suitable for attributing and mapping census metrics at the district and taluk level.

Observing the models for all the three levels of administrative regions, it was noted that at the districts, 50% of the proposed models used only brightness information, 30% used only stable lights data and 20% used both brightness and stable lights. At the taluks, 66% of the optimum models used brightness as predictor while 22% used only stable lights and 12% used both brightness and stable lights as obtained from the images. At the villages, only brightness information was used to propose the models. From these observations, it was inferred that DMSP-OLS image with brightness information was the most suitable in predicting the models which answered research question 3 (section 1.2). The stable light images record consistent lights from settlements while suffering from saturation at the bright areas. The problems with the images were overcome in the radiance calibrated images of 2006 (Ziskin et al., 2010).

7. Applications of the models to produce maps of non-available census metrics for small regions

This chapter describes the process of deriving maps of census metrics not collected by Indian census for small regions using DMSP-OLS images that are otherwise unavailable. Maps are produced for villages and for areas as small as one square kilometre using the models proposed in chapter 6. This chapter describes errors associated with the Modifiable Areal Unit Problem (MAUP) and ecological fallacy. This is followed by description of maps for the villages and at one square kilometre areal units, and concludes with an overall assessment of the results at these various spatial scales.

7.1: Introduction:

From the ten census metrics selected for this study (section 4.8.1), only three variables were available from the Indian census at the scale of a village. They are *number of households per square kilometre*, *total population per square kilometre* and *total workers per square kilometre*. This chapter examines the application of the models proposed in chapter 6 to predict and map the metrics for the villages and smaller areas that are unavailable from traditional census statistics. These include: *number of female literates per square kilometre*; *percentage of households with cars, jeeps and vans*; *percentage of households with television*; *percentage of permanent census houses* and *percentage of households using electricity as power source*. Maps were produced for the villages of Pune, Yavatmal and Gadchiroli. The use of multi-scale data led to the consideration of issues arising from MAUP and ecological fallacy which are also described in this chapter.

7.2: Issues surrounding multi-scale data analyses:

The effect of scale on statistical results was first demonstrated by Gehlke and Biehl (1934, as cited in Dark and Bram, 2007, Dungan et al., 2002, Openshaw, 1984) and Yule and Kendall (1950 cited in Marceau, 1999, Marceau and Hay, 1999). McCarthy (1956 cited in Marceau, 1999) showed that the statistical results valid at one spatial scale may not be applicable at another scale. The problems of scale differences can be described as the determination of appropriate spatial scale to study a particular geographical phenomenon, and the transferability of information between two spatial scales (Marceau, 1999). The significant effect of spatial aggregation of data was acknowledged by the pioneer works of Blalock (1964) Clark and Avery (1976) and Fotheringham (1991). The most common errors arising from the use of multi-scale data are MAUP (Openshaw, 1984, Doll et al., 2004, Marceau, 1999) and ecological fallacy (Cao and Lam, 1997, Doll et al., 2004, Robinson, 1950).

The MAUP can affect the results in spatial studies using aggregate data sources (Unwin, 1996). Some geographical data are aggregated over arbitrary boundaries such as census collection districts and postcode zones. The boundaries of these zones are "modifiable" in nature and is often "… *subject to the whims and fancies of whoever is doing, or did, the aggregating*" (Openshaw, 1984). The MAUP consists of two components: the scale effect and the aggregation effect (Doll et al., 2004, Marceau, 1999, Marceau and Hay, 1999). The scale effect is observed when data from small regions are aggregated into larger spatial units (Doll et al., 2004, Wrigley et al., 1996). Aggregation effect takes place due to the combining of zone boundaries in a given scale of analysis (Doll et al., 2004).

The effects of scale and aggregation are usually manifested in several ways in studies in spatial analyses depending on the generalization of the datasets. The scale effect is demonstrated through individualistic fallacy and ecological fallacy, while the zoning or aggregation effect gives rise to cross - level fallacy. Individualistic fallacy occurs when the inferences from small or micro - levels are used to infer results for macro regions. Ecological fallacy can be regarded as the opposite of individualistic fallacy and is observed when inferences about micro - regions are derived from relationships at macro - regions (Cao and Lam, 1997, Doll et al., 2004). Cross - level fallacies are found in inferences derived for one sub - population from another at the same spatial scale of analysis (Doll et al., 2004).

There are many different approaches proposed in the literature for managing issues of MAUP (Fotheringham, 1989, Openshaw, 1984, Marceau, 1999, Marceau and Hay, 1999). For example, Openshaw (1984) proposed the approach of optimal zoning system for spatial analyses. An optimum scale was defined as "… *the spatial sampling unit corresponding to the scale and aggregation level characteristic of the geographical entity of interest*" (Marceau and Hay, 1999). An important consideration of optimal scale approach was the absence of unique optimal resolution. Another approach to manage MAUP was the identification of basic entities. This approach necessitated the study of an object of concern at a spatial scale where it could be observed and measured (Fotheringham, 1989, Visvalingam, 1991). The object was aggregated in the entity based approach and therefore this was one of the most effective ways to overcome MAUP (Fotheringham, 1989).
Commonly used ones include abandonment of traditional statistical analyses and sensitivity analyses.

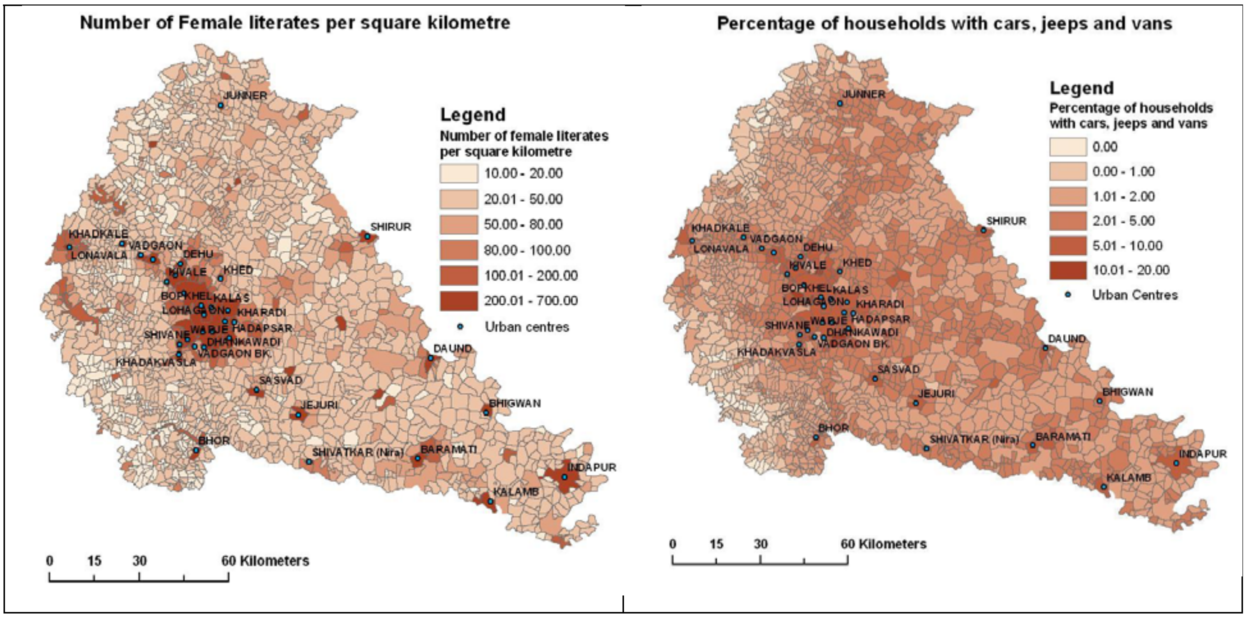
The use of traditional statistics is limited in of its application to spatial data. Recent studies in remote sensing indicated the use of spatial statistics such as geo-statistical tools and autocorrelation indices in order to overcome the effect of MAUP (Marceau and Hay, 1999).

In order to incur the least possible error from MAUP, this study used the approach of optimal zoning system. In the Indian census, the villages are aggregated to form taluks. Therefore the taluks were considered to be the optimal aggregation unit to propose the metrics for the villages. The optimal models for the taluks (Table 6B.5) were used to predict the metrics for the villages. The results from the predicted metrics were mapped for the districts of Pune, Yavatmal and Gadchiroli, the three districts used to propose models for villages (Table 6C.3). The predicted metrics are presented in the following sections.

7.3: Results and discussion

7.3.1: Maps for villages

Number of female literates per square kilometre; *percentage of households with cars, jeeps and vans*; *percentage of households with television*; *percentage of permanent census houses* and *percentage of households using electricity as power source* were predicted and mapped at the village level. The maps with the predicted metrics for the districts of Pune, Yavatmal and Gadchiroli are shown from figures 7.1 to 7.3.



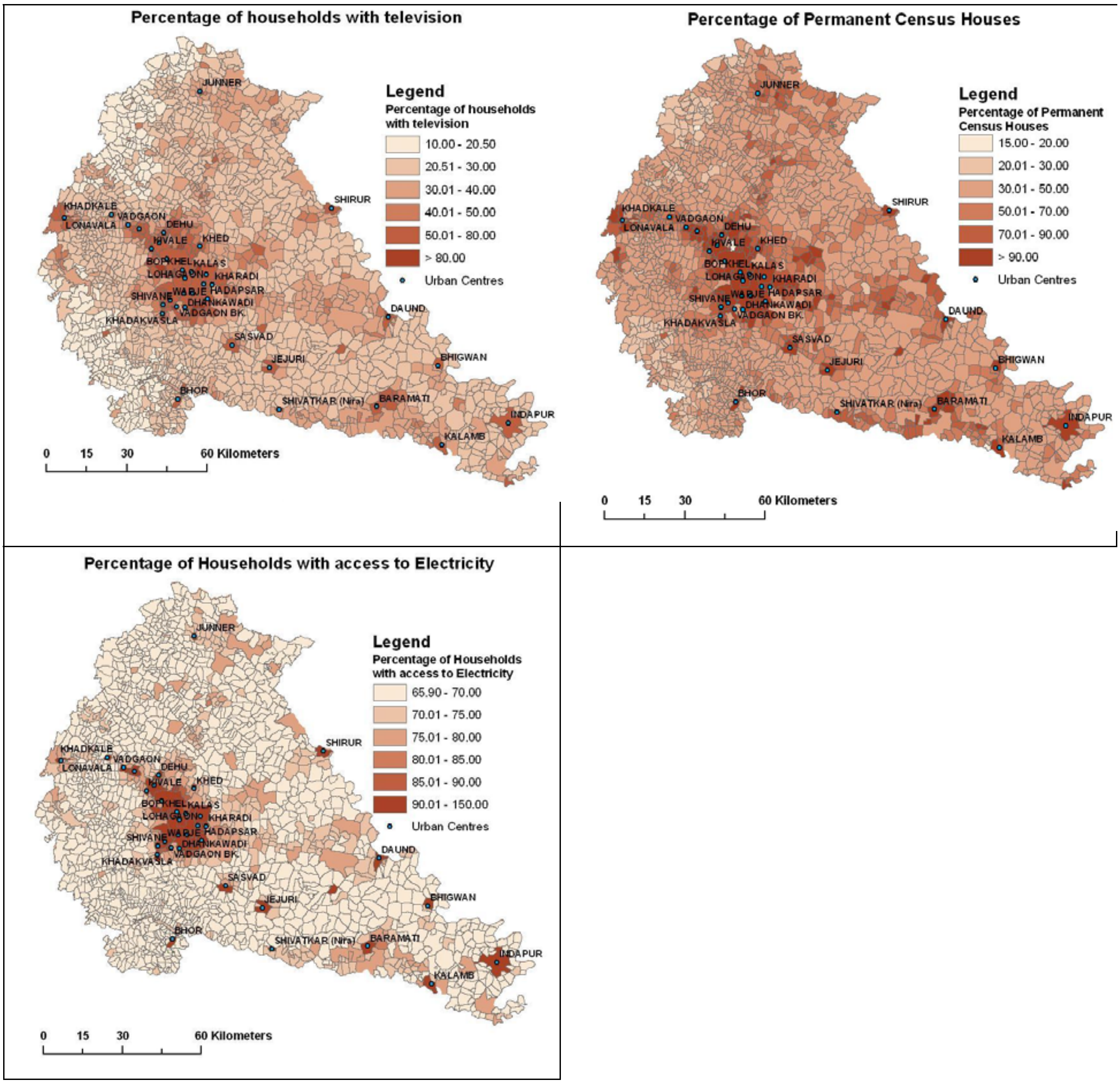


Figure 7.1 : Predicted maps of metrics not collected by the census for villages of Pune

In Pune, high values of number of female literates were predicted for villages around the urban areas such as Vadgaon Bk, Hadapsar, Khed, Kharadi, Kivale and Dehu in the central part of the district and Jumner, Shirur, Baramati, Kalamb and Bhor in other parts of the district. These areas were predicted to have more than 200 *female literates per square kilometre*. Most of the villages in the district have approximately 20 to 80 *female literates per square kilometre*. On an average two to five percent of the households in the villages were predicted to have cars, jeeps and vans. Around the urban centres there were five to ten percent of the households predicted as having cars, jeeps and vans. Similar trends were predicted for *percentage of permanent census houses*. Urban areas were recorded to contain 70% to more than 90 % of permanent houses. The villages in the district as a whole showed to have 30 to 50 % of permanent houses. More than 85% of the households demonstrated to have access to electricity near the urban areas while the district as a whole exhibited 65% to 70% of rural electrification.

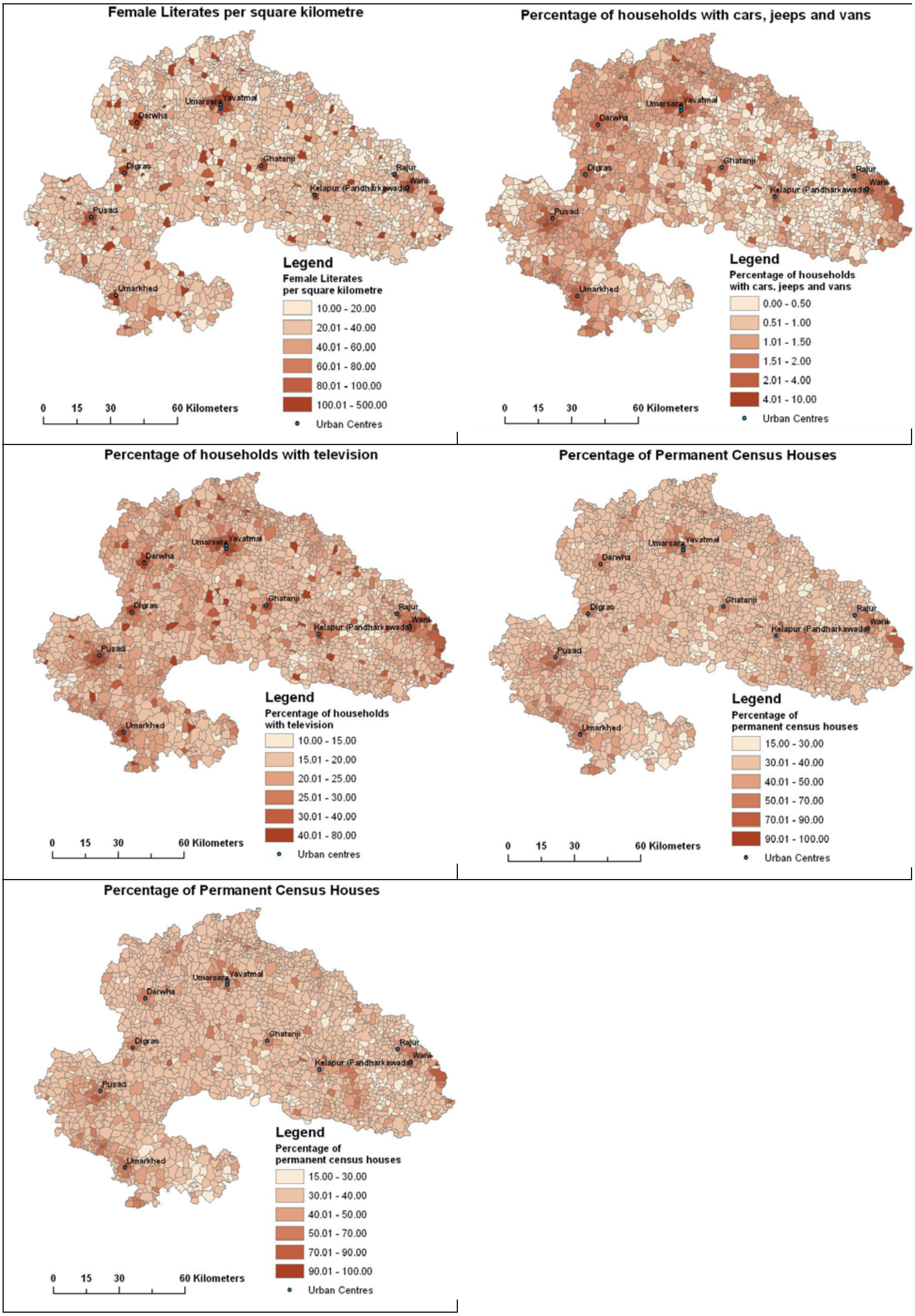
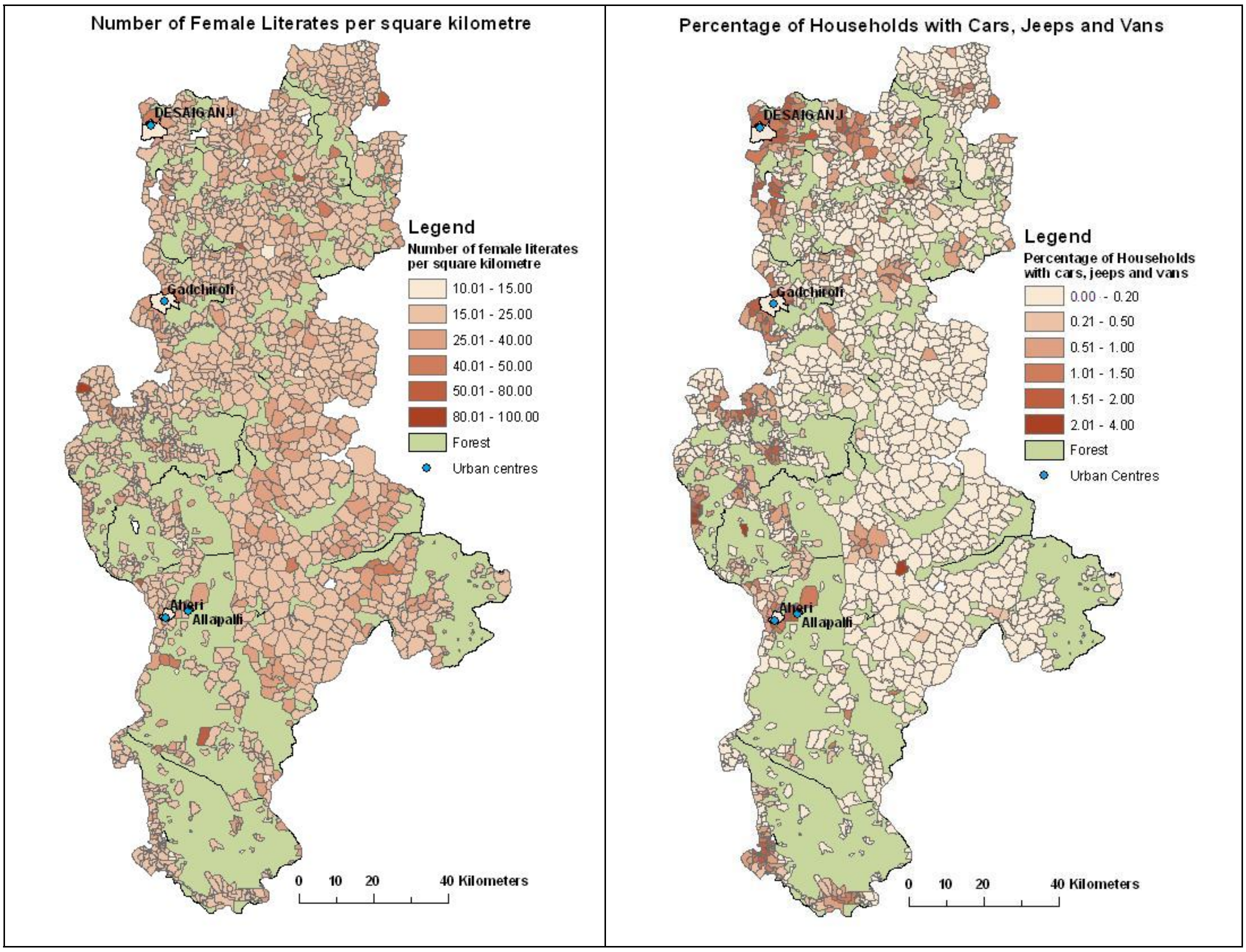


Figure 7. 2 Predicted maps of metrics not collected by the census for villages of Yavatmal

In Yavatmal, the villages on an average were predicted to have 40 to 60 *female literates per square kilometre*. The villages around the urban centres of Yavatmal and Umarsara, however, exhibited high predicted values of around 100 to 150 *female literates per square kilometre*. *Cars, jeeps and vans* were used by 1.5 to 2 % of households all over the district. Areas around the urban centres of Wani, Umarkhed, Pusad, Digras, Yavatmal and Umarsara were predicted with more than 4% of *households with cars, jeeps and vans*. The figures show higher accessibility of these villages to the towns. 40 to 50 % of the census houses in Yavatmal were permanent in nature with 40 to 50% of the households in the villages with access to electricity.



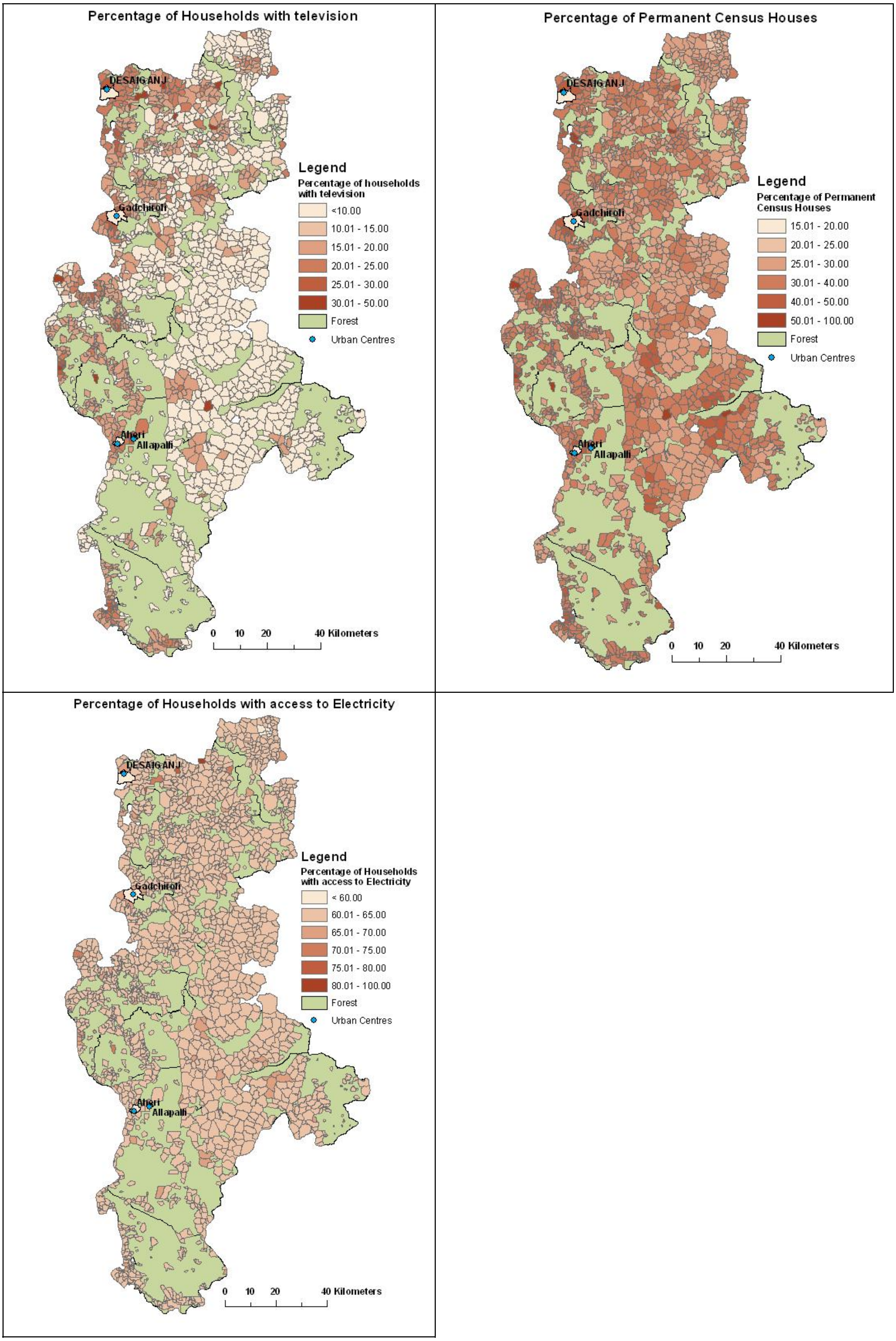


Figure 7.3 : Predicted maps of metrics not collected by the census for villages of Gadchiroli

The district of Gadchiroli is different from both Pune and Yavatmal in that there is a vast tract of forest area of more than 5000 square kilometre in the district. The forest extends mainly across the south western part of the district with some areas extending to the eastern and northern part. Most of the villages in the district were predicted with 15 to 25 *female literates per square kilometre*. In this district, less than 1% of the households in most of the villages were predicted to have cars, jeeps and vans. Areas around the towns of Desaiganj, Gadchiroli and Aheri in the western part only exhibited to around two to four percent of the *households with cars, jeeps and vans*. Although 60 to 65% of the villages were predicted to have access to electricity, the villages around the urban areas demonstrated to have more than 20% of the *households with television*. Less than 10% of the households in the rest of the district were predicted to use television. Most of the villages in Gadchiroli had more than 40% of the census houses as permanent.

Due to the absence of these metrics at the village level from the census, it was not possible to validate the results obtained from the predictions. The results presented in this chapter are the best estimates of the metrics previously unavailable.

7.3.2: Maps for smaller regions

One of the novel applications of the models proposed in this research was to predict census metrics for small areas. The brightness and stable light information obtained from the images were used to predict metrics and produce maps for areas as small as one square kilometre. This is the spatial resolution of the DMSP-OLS night-time images. The maps of *number of households per square kilometre*, *total population per square kilometre* and *total workers per square kilometre* for part of Pune are shown in figure 7.4. These metrics were predicted using the models already validated over the villages in chapter 6C (table 6C.3). The approach of optimal zoning system was also used in this case to overcome the effects of MAUP. A village ranges in area from a few square kilometres to a few hundreds of square kilometres. Therefore, they were considered to be the most suitable aggregation unit to predict metrics for one square kilometre areal units.

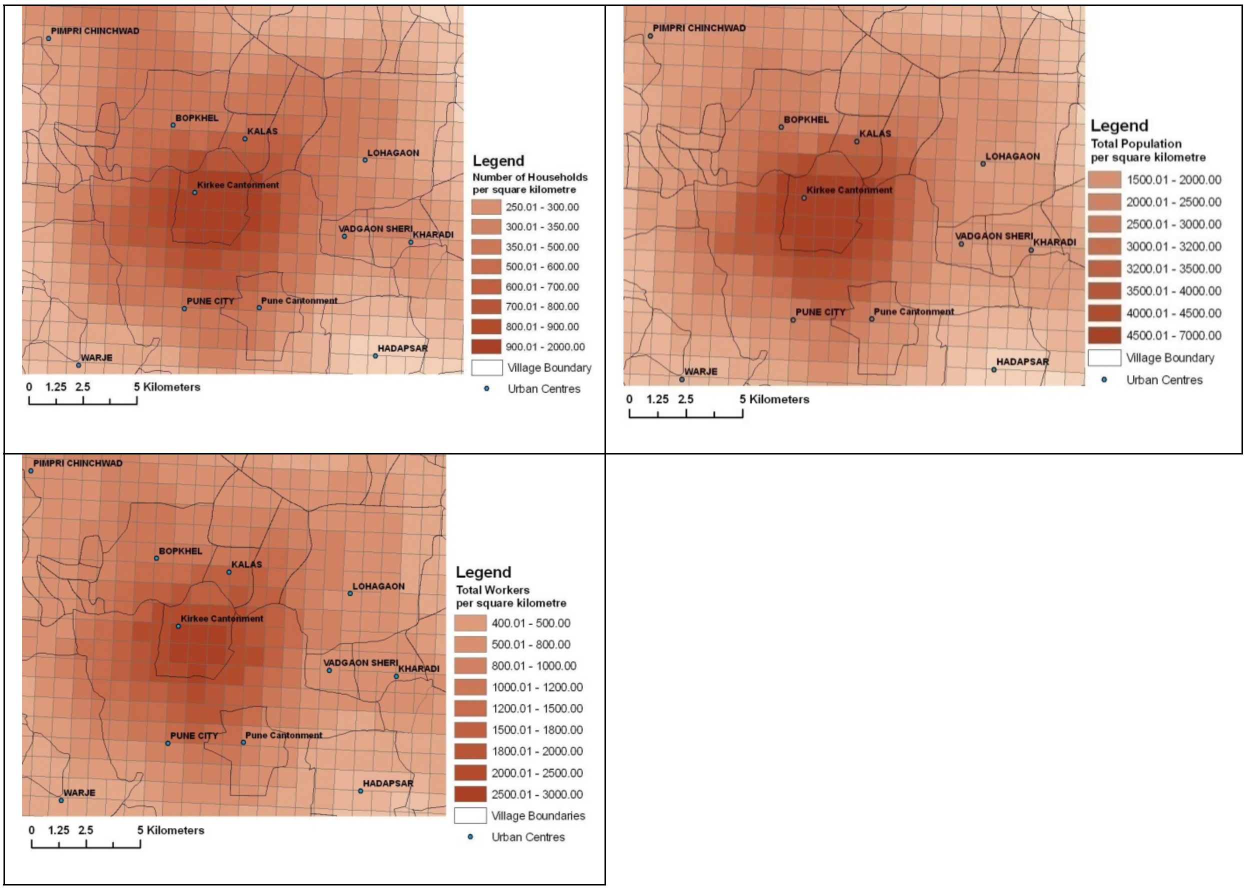


Figure 7.4 : Predicted maps of metrics not collected by the census for one square kilometre areas in the villages of Pune

The maps show variations in the census metrics within a village. It was observed that areas around Pune city and Kirkee Cantonment had more than 600 *households per square kilometre*. These areas also demonstrated the highest concentration of total population. *Population density* in parts of these areas ranged from 3500 to 7000 *people per square kilometre*. The working population in the areas around the urban centres varied from 1500 to 3000 per square kilometre. From the maps it was observed that the distribution of the metrics became less dense with distance further from the urban centre. For example, compared to the centre of Pune city, in Pune Cantonment area, the *number of households per square kilometre* ranged from less than 250 to 500 with a *population density* from 1500 to 2500 *persons per square kilometre*. Further variation in the range of values was noted in areas surrounding Lohagaon and Hadapsar. These areas were less densely populated than the urban core.

7.3.3: Comparison of maps from all spatial scales:

Census metrics were proposed and predicted over different spatial scales in this study. Chapters 5 and 6 used the DMSP-OLS night-time images to propose models for a surrogate census. This chapter used the proposed models in predicting the unavailable census metrics. It also showed how the models can be effectively used to propose metrics for small areas within administrative boundaries. Maps of *number of households per square kilometre* in all the spatial scales are shown in figure 7.5. From the map at the district level (figure 7.5 (a)) it was observed that in Pune, as a whole, there were 40 to 50 households per square kilometre. A look at smaller administrative areas such as taluks and villages, revealed detailed distribution of the households. The map of *number of households per square kilometre* at the taluk level is shown in figure 7.5 (b). It was observed that taluks in the north western, eastern and southern part of Pune district exhibited the lowest density values of around 20 to 50 households per square kilometre. These taluks comprised of an area of around 7000 square kilometre in the district. Five taluks demonstrated to have moderate density of 80 to 200 *households per square kilometre*. These taluks occupied an area of approximately 6000 square kilometre in Pune. Only 1400 square kilometre of the district had high values of 300 to 500 *households per square kilometre*. These taluks were located near the centre of the district around the urban areas of Pune. Since most of the taluks exhibited to have low to moderate household density, the aggregated district map showed a low value for the district as a whole.

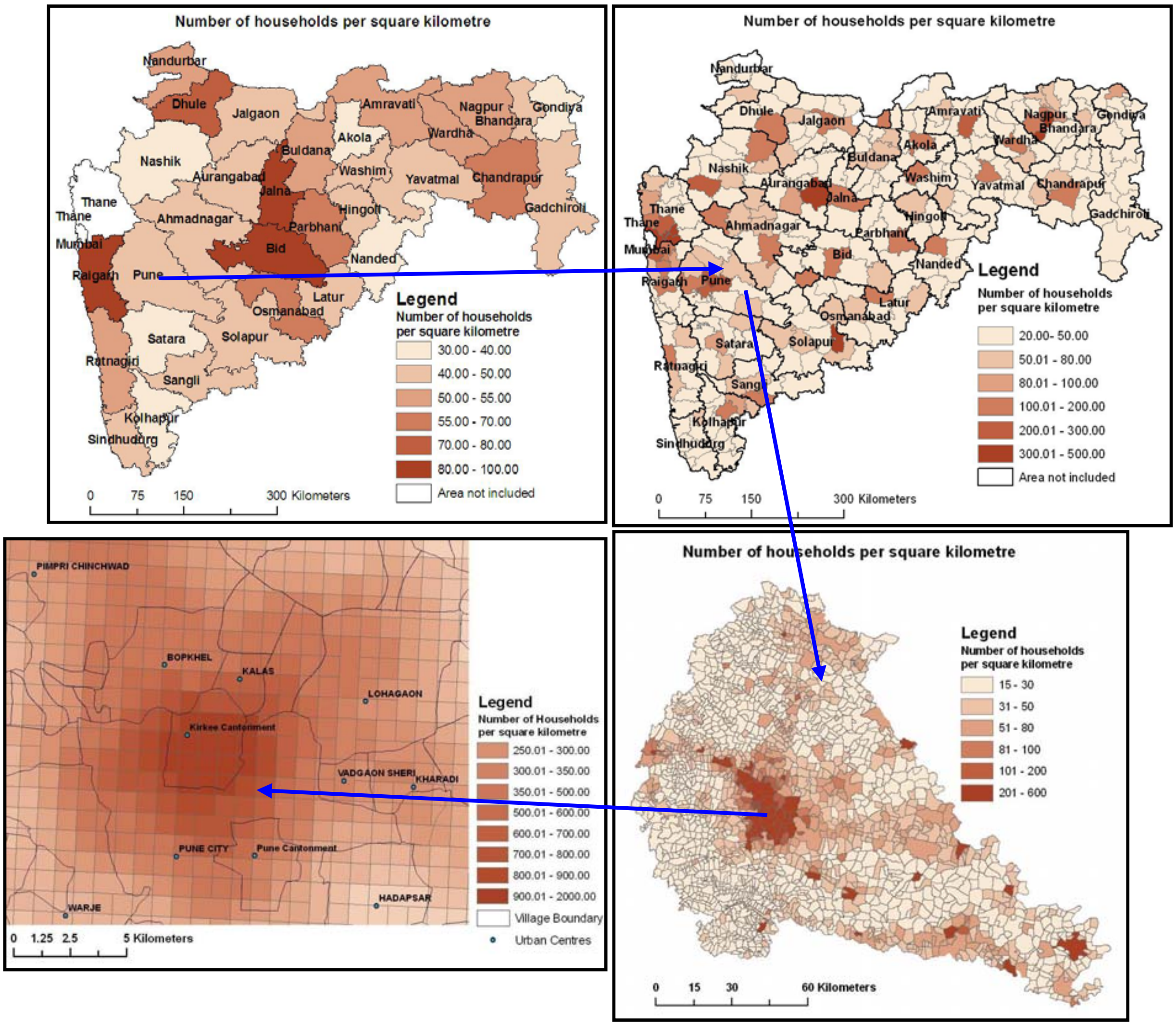


Figure 7.5 : The effect of scale in the prediction of census metrics at (a) Districts; (b) Taluks; (c) Villages and (d) one square kilometre areas.

of the district around cities exhibited highest household densities of more than 100 *households per square kilometre*. Moderate distribution (50 to 100 *households per square kilometre*) was noted in the north eastern and southern part while the major part of the area had around 15 to 30 households per square kilometre. Therefore it was apparent from these maps that values of metrics aggregated over small areas influenced the data of large regions. The analyses of the results in all the spatial scales helped overcome the individualistic and ecological fallacies.

7.4: Summary:

This chapter looked into the application of the models for proposing census metrics otherwise not collected by the census at the village level and for one square kilometre areal units. The errors arising from analyses of multi – scale data such as MAUP and ecological fallacy were examined and the approach of optimal zoning system was used to overcome the MAUP effects in predicting the metrics for the villages of Pune, Yavatmal and Gadchiroli. However, due to the absence of these data from the census, it was not possible to validate the results. This chapter showed the potential for models derived from DMSP-OLS images for mapping and predicting census metrics for small regional scales.

Key Findings from the Chapter:

- The models proposed in chapter 6 were used to predict census metrics not enumerated in the official Indian census.
- The approach of optimal zoning system was used to overcome the problem arising from MAUP.
- DMSP-OLS images exhibited the potential for mapping and predicting census metrics for small regions.

8. Conclusion:

8.1: Summary of Research:

This study commenced with a review of the relevant literature on the application of remote sensing, especially DMSP-OLS images, in socio-economic and census studies and is presented in Chapter two. For this research, the census is considered a major database on socio-economic and demographic parameters for India. It is widely used for various planning and research purposes and typically consists of metrics such as population composition, literacy levels and data on working population. Remote sensing offers many opportunities for census mapping and estimation and is a valuable and complementary tool along with the traditional method of census estimation for improving the accuracy and currency of traditional censuses. The DMSP-OLS derived dataset is presently the only dataset available that records light emitted from cities at night in the visible band of the electro-magnetic spectrum. This data has demonstrated application in the areas of urbanization, economic development, environmental conditions, species' behaviour and disaster management and has been used to produce new datasets from global poverty maps to global maps of impervious surfaces.

The study area for this research was the state of Maharashtra in western India and is described in chapter three. Chapter four presented the processing and preliminary analyses of census data and satellite images. It also describes the methods undertaken to prepare the census data and the satellite images for further statistical and spatial analyses. A list of 10 census metrics was shortlisted to develop models using fixed gain radiance calibrated datasets as well as global composite data on brightness and stable lights. There were a number of shortcomings associated with these datasets and these are also discussed in chapter four. Following treatment of the limitations of the data, models were then developed and validated to test the utility of fixed gain datasets and also to determine the optimum gain needed to propose the surrogate census metrics.

Chapter five used the single orbit fixed gain radiance calibrated DMSP-OLS datasets to propose surrogate census at the district level. It was found that at the district level, the gain 50 dB image was more useful than the gain 20 dB image for proposing the census. The global composite images were also analyzed and models proposed for three spatial scales. The results for this dataset are given in chapter six. The application and the utility of the models from the global composites of stable lights and brightness datasets are described in chapter seven. In this chapter, the models were used to predict census metrics for small regions such as villages for which these metrics are typically not collected or reported.

8.2: Key Findings

This research proposed a method that enables the prediction of census metrics more frequently than is currently available. The ability to estimate census metrics at frequent time intervals will assist in population policy making and development planning activities. Another benefit of this research is that it enables the prediction of otherwise unavailable census metrics at the sub-district level as well as areas as small as approximately one square kilometre (i.e. the area represented by one pixel in the satellite images used). The approach proposed in this research can be applied independent of pre-defined or administration boundary systems thereby allowing for census summary statistics to be predicted for any user defined area, administration unit or boundary.

This thesis presented three main research questions necessary to understand the opportunities and limitations the DMSP-OLS datasets had for informing the census in India. The key findings of each research question are summarized as follows.

RQ1: Which socio-economic metrics obtained from the census can be correlated with light information obtained from DMSP-OLS satellite images?

The census and satellite datasets were processed and analyzed to determine which available census metrics can be correlated with information on light obtained from DMSP-OLS images. The analyses started with 144 variables. Different statistical tests were conducted on the available census metrics for determining which socio-economic metrics obtained from the census could be correlated with light information obtained from the DMSP-OLS satellite images. The statistical tests included the process of sampling, tests for normal distribution and bootstrapping the correlation coefficients. From these tests a shortlist of census metrics was prepared by the process of elimination.

The districts in the state of Maharashtra were randomly chosen for model development. There are 35 districts in the state. Of these, the districts of Mumbai, Greater Mumbai and Thane were removed from the study because of very high values of brightness and stable lights compared to others. From the remaining 32 districts, 24 were selected randomly and 8 districts were withheld for validating the models. The tests for normal distribution included plotting of histograms and normal – probability plots, test of skewness and kurtosis and analyzing the goodness of fit by Shapiro and Wilks test. In order to test the significant metrics, mean and standard deviation of radiance, brightness and stable lights were obtained from the images and used for calculating correlations using the method of bootstrapping. Bias and standard error of all the correlations was noted. Variables with a bias of less than 0.05 and standard error of less than 0.2 were chosen. The results were used to select the final variables used in this research. The selected variables were:

- *Number of households per square kilometres*
- *Total population density*
- *Urban population density*
- *Female literates per square kilometres*
- *Total number of workers per square kilometres*
- *Percentage of households with car, jeep and van*
- *Percentage of households with access to electricity as power source*
- *Percentage of households with television*
- *Percentage of permanent houses*
- *Per Capita District Domestic Product*

RQ2: What is the most appropriate spatial resolution or mapping unit for attributing the census metrics using DMSP-OLS night-time images?

This was addressed by comparing the significant r^2 values from the correlations between all metrics at the three spatial scales (that is, villages, taluks and districts), and the mean and standard deviation of stable lights and brightness information as obtained from satellite images. The performances of the proposed models at the three spatial scales were also explored.

Comparison of the results at all the three spatial scales (as obtained from chapters 5 and 6), it was observed that at the scale of districts, the correlation coefficients (r) ranged from 0.3 to 0.9 at the 95% confidence interval. At the taluk level, the correlations ranged from 0.3 to 0.8 at the 95% confidence interval. All the chosen census metrics exhibited significant correlations at the taluk level. For villages, however, there were significant correlations between census metrics and means of brightness and stable lights only at the 90% confidence interval. The values of the correlation coefficients were below 0.3 for the available census metrics at the spatial scale of villages. Overall it was observed that the correlations were higher for the districts than for the taluks.

More than 75% of the districts were predicted within a 25% error margin by the models proposed from fixed gain images. Variables such as *Number of households per square kilometre* and number of *Total workers per square kilometre* were predicted optimally for 84% of the districts. *Population density, number of households with access to electricity, PCDDP* and *percentage of households with television* were predicted for more than 75% of the districts. However, *female literates per square kilometre* and *percentage of households with cars, jeeps and vans* were optimally predicted for 41% and 47% of the districts respectively. However, the data suffers from a number of problems such as from cloud cover and fires. Therefore, these data were not used for further analyses in this research for predicting census metrics at small spatial scales (taluks and villages). Using the global composite images, at the district level, most of the models performed with more than 70% accuracy for most of the metrics. *PCDDP* was predicted optimally for 87.5% of the districts followed by 84.4% for *number of households per square kilometre*. However, 56% of the districts were predicted within the error margin of |25|% for *percentage of households with cars, jeeps and vans*. At the taluk level, success in the predicted census variables using the models varied from 60% to 20%. Unlike the models at the district level, most of the census parameters could be predicted with one optimum model at the taluk level. For the villages, the models predicted all the metrics with moderate accuracy, with around 27% to 38% of the villages being predicted within the |25|% error margin.

Therefore it was concluded that the most appropriate spatial resolution or mapping unit for attributing the census metrics using DMSP-OLS night-time images was at the district level. The second best level for application of the models to map census metrics was at the spatial scale of the taluks. However, unlike larger areas such as districts where DMSP-OLS night-time images adequately predict census metrics, at the sub-district level such as taluks, the results need to be supplemented and validated with other information sources such as survey reports. These findings demonstrate that the DMSP-OLS images are suitable for attributing and mapping census metrics at the district and taluk level.

RQ3: Which night-time satellite image product obtained from the DMSP-OLS satellite is most suitable for the purpose of mapping surrogate census metrics?

In order to answer this question, models were proposed and adjusted r^2 from the models were compared for all regional scales separately. In chapter five, the single orbit fixed gain radiance calibrated datasets were used to test the utility of these datasets in proposing surrogate census. Observing the models for all the three levels of administrative regions from chapter six, it was noted that at the district level, 50% of the proposed models used only brightness information, 30% used only stable lights data and 20% used both brightness and stable lights. At the taluk level, 66% of the optimum models used brightness as predictor while 22% used only stable lights and 12% used both brightness and stable lights as obtained from the images. At the village level, only brightness information was used to propose the models. From these observations, it was inferred that DMSP-OLS image with brightness information was the most suitable in predicting the models.

8.3: Limitations:

This research was limited to the availability of the census datasets and the administrative boundaries datasets as provided by the SAC, India. Inconsistencies existed between these two datasets. For example, in some areas there was a lack of positional agreement or alignment between the collection and administrative boundaries. Given this, the study was restricted to only those taluks and villages for which spatial data was available from the SAC. In addition, the absence of an inter-censal database on socio-demographic metrics in India meant this research had to use census data collected from 2001. As a result, global composite data on brightness information from DMSP-OLS satellite had to be used in order to be temporally comparable.

Another limitation faced in this research was the absence of census data at the level of taluks and villages. For example, *PCDDP* or any metrics on productivity was not available from the census at the scale of taluks and villages. Therefore, at the Taluk level the models were predicted using: *number of households per square kilometre, total population per square kilometre, female literates per square kilometre, total workers per square kilometre, percentage of households with cars, jeeps and vans, percentage of households with television, percentage of permanent houses, percentage of households using electricity as power source* and *urban population per square kilometre* (described in section 6B). At the village level, data on only three demographic variables were available from the census. These were: *number of households, total population* and *total number of workers* (described in section 6C). As a result, the development of the models was limited by the availability of the census metrics. Despite this limitation, the unavailable metrics were predicted at the village level and also for areas as small as 1 square kilometre (described in chapter 7). However, due to the absence of these data in the census, the model outputs could not be validated.

The availability of DMSP-OLS fixed gain datasets is limited both in terms of their numbers and temporal frequency because there were only a few days of the year when the sensor flew with a fixed gain. On the basis of the study area chosen (the state of Maharashtra in this case), two fixed gain images, one with gain setting 20dB and the other with gain 50 dB were chosen from DMSP satellite F15. The G20 image was captured on 17th September, 2001 and the G50 image was collected on 14th September of the same year. Using the reference datasets available at Space Physics Interactive Data Resources (SPIDR) and the DMSP-OLS thermal infra red images, dates with no cloud cover over the study area were selected. Due to this limitation, it was assumed in the study that the radiances recorded on the chosen nights were typical snapshots of the radiance recorded over the study area over a whole year period. In addition to this, the fixed gain data suffers from a number of problems such as from fires and the lights tend to vary in brightness from night-to-night As a result, the single orbit fixed gain radiance calibrated datasets were not used for predicting census metrics at the spatial scales of taluks and villages

8.4: Recommendations for Future Research:

This research demonstrates the potential for DMSP-OLS data to be used to support and provide census collection data. However, further research is suggested and can be considered under two broad categories: overcoming the limitations faced in this current research and future applications of the data.

As identified in section 8.3, a limitation of the research was the absence of important census metrics at spatial scales lower than districts. The availability of the census metrics at all the spatial scales in future would facilitate improvements in model development and validation. Overcoming the problem of inconsistency between spatial datasets and census data would assist in a higher spatial agreement that is needed for accurate merging of spatial and attribute datasets. This would then enable easy and widespread application, interpretation and map production for all administrative regions. Frequent availability of high and low fixed gain datasets would aid further research in policy and development planning by using those datasets. Monthly DMSP-OLS datasets are now provided by NGDC and are showing utility for temporal change detection of socio-economic metrics over regions of varying spatial scales.

This research has shown the independence and flexibility of the DMSP-OLS derived models to be used for predicting census metrics at different scales of administration zones. The utility of these models may be further developed by applying them to alternative regions from those defined by administration boundaries such as planning regions, or even biophysical zones. More research is still required to fully understand the limitations and reliability of DMSP-OLS derived models for estimating and monitoring census metrics in these alternative land zones. The models themselves could even be adapted for delineating urban, sub–urban and rural areas thereby developing new zoning systems and identifying new settlement growths over time. In particular, variations in light intensity as recorded by the DMSP-OLS images may offer benefit in measuring resource allocation and infrastructure distribution at sub–national levels. Overall, DMSP-OLS datasets and derived models have the potential to be widely used in areas planning and regional development and, as such, further research is needed to achieve this potential.

9. Bibliography

Amaral, S., Camara, G., Monteiro, A. M. V., Quintanilha, J. A. & Elvidge, C. D. 2005. Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. *Computers, Environment and Urban Systems,* 29**,** 179-195.

Amaral, S., Monteiro, A. M. V., Camara, G. & Quintanilha, J. A. 2006. DMSP/OLS nighttime light imagery for urban population estimates in the Brazilian Amazon. *International Journal of Remote Sensing,* 27**,** 855-870.

Anderson, B. 1999. *Imagined communities: Reflections on the origin and spread of nationalism,* London, Verso.

Anderson, S. J., Tuttle, B. T., Powell, R. L. & Sutton, P. C. 2010. Characterizing relationships between population density and nighttime imagery for Denver, Colorado: Issues of scale and representation. *International Journal of Remote Sensing,* 31**,** 5733-5746.

Armston, J. D., Danaher, T., Goulevitch, B. M. & Byrne, M. I. 2002. Geometric correction of Landsat MSS, TM, and ETM+ imagery for mapping of woody vegetation cover and change detection in Queensland. *Proceedings of the 11th Australasian Remote Sensing and Photogrammetry Conference.* Brisbane, Australia.

Badarinath, K. V. S., Sharma, A. R. & Kharol, S. K. 2011. Forest fire monitoring and burnt area mapping using satellite data: a study over the forest region of Kerala State, India. *International Journal of Remote Sensing,* 32**,** 85-102.

Balk, D., Pozzi, F., Yetman, G., Deichmann, U. & Nelson, A. 2005. The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents. *International Society for Photogrammetry and Remote Sensing, Proceedings of the Urban Remote Sensing Conference, March*.

Bartlett, J. E., Kotrlik, J. W. & Higgins, C. C. 2001. Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning and Performance,* 19**,** 43 - 50.

Baugh, K., Elvidge, C. D., Ghosh, T. & Ziskin, D. 2010. Development of a 2009 Stable lights Product using DMSP - OLS data. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

BBC News. 2011. India census: population goes up to 1.21bn. *BBC News South Asia* *http://www.bbc.co.uk/news/world-south-asia-12916888*, 31st March, 2011.

Beaud, J. P. & Prévost, J. G. 2005. Statistics as the science of government: The stillborn British Empire statistical bureau, 1918–20. *The Journal of Imperial and Commonwealth History,* 33**,** 369-391.

Bhaduri, B., Bright, E., Coleman, P. & Dobson, J. 2002. LandScan: Locating people is what matters. *Geoinformatics,* 5**,** 34-37.

Bhatta, B. 2008. *Remote Sensing and GIS,* India, Oxford University Press.

Blalock, H. M. 1964. *Causal inferences in nonexperimental research,* University of North Carolina Press.

Brugioni, D. A. 1983. The census: it can be done more accurately with space-age technology. *Photogrammetric Engineering and Remote Sensing,* 49**,** 1337-1339.

Cao, C. & Lam, N. S. N. 1997. Understanding the scale and resolution effects in remote sensing and GIS. *Scale in Remote Sensing and GIS.* Florida: CRC Lewis.

Cao, X., Chen, J., Imura, H. & Higashi, O. 2009. A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data. *Remote Sensing of Environment,* 113**,** 2205-2209.

Central Statistical Office. 2009. *National Accounts Statistics: Manual on Estimation of State and District Income* [Online]. Available: http://www.mospi.gov.in/national_accounta.pdf [Accessed 20th November, 2010].

Chand, T. R. K., Badarinath, K. V. S., Elvidge, C. D. & Tuttle, B. T. 2009. Spatial characterization of electrical power consumption patterns over India using temporal DMSP-OLS night-time satellite data. *International Journal of Remote Sensing,* 30**,** 647 - 661.

Chapman, A. D. 2004. *Principles of data quality,* Copenhagen, Report for the Global Biodiversity Information Facility.

Chen, K. 2002. An approach to linking remotely sensed data and areal census data. *International Journal of Remote Sensing,* 23**,** 37 - 48.

Chernick, M. R. 2008a. The Bootstrap Method. *Bootstrap methods: A guide for practitioners and researchers.* Wiley-Interscience.

Chernick, M. R. 2008b. What is bootstrapping? *Bootstrap methods: A guide for practitioners and researchers.* Wiley-Interscience.

Christopher, A. J. 2008. The quest for a census of the British Empire c.1840-1940. *Journal of Historical Geography,* 34**,** 268-285.

Cinzano, P., Falchi, F. & Elvidge, C. 2007. Recent Progress on a second world atlas of the night sky brightness:LPTRAN/LPDART realistic models,tomography of light pollution, accurate validation methods and extended satellite data analysis. *Starlight 2007.* Istituto di Scienza e Tecnologia dell'Inquinamento Luminoso (ISTIL). Italy.

Cinzano, P., Falchi, F. & Elvidge, C. D. 2001. The first World Atlas of the artificial night sky brightness. *Monthly Notices of the Royal Astronomical Society,* 328**,** 689-707.

Clark, W. A. V. & Avery, K. L. 1976. The effects of data aggregation in statistical analysis. *Geographical Analysis,* 8**,** 428-438.

Clarke, D. G. & Clark, D. M. 1995. Lineage. *In:* Guptill, S. C. & Morrison, J. L. (eds.) *Elements of Spatial Data Quality.* Oxford, UK: Elsevier Science, published on behalf of the International Cartographic Association.

Cochran, W. G. 2009. *Sampling techniques*, Wiley-India.

Comber, A. J., Fisher, P. F., Harvey, F., Gahegan, M. & Wadsworth, R. 2006. Using Metadata to Link Uncertainty and Data Quality Assessments
In: Riedl, A., Kainz, W. & Elmes, G. A., eds. Progress in Spatial Data Handling 12th International Symposium of Spatial Data Handling, Springer, Berlin.

Coulter, L. L. & Stow, D. A. 2009. Monitoring habitat preserves in southern California using high spatial resolution multispectral imagery. *Environmental Monitoring and Assessment,* 152**,** 343-356.

Croft, T. A. 1978. Nighttime images of the earth from space. *Scientific American,* 239**,** 86-96.

Dark, S. J. & Bram, D. 2007. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography,* 31**,** 471-479.

de Sherbinin, A., Balk, D., Yager, K., Jaiteh, M., Pozzi, F., Giri, C. & Wannebo, A. 2002. *Social Science Applications of remote sensing* [Online]. Available: http://sedac.ciesin.columbia.edu/tg/guide_frame.jsp?rd=RS&ds=1 [Accessed 12th June, 2008].

Desrosières, A., Naish, C. & Hunter, P. W. 2006. The politics of large numbers. A history of statistical reasoning. *The Mathematical Intelligencer,* 28**,** 69-70.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. & Worley, B. A. 2000. LandScan: A Global Population database for estimating populations at risk. *Photogrammetric engineering and remote sensing,* 66**,** 849-857.

Doll, C. N. H. 2003. Estimating non-population activities from night-time satellite imagery. *In:* Mesev, V. (ed.) *Remotely Sensed Cities.* London: Taylor and Francis.

Doll, C. N. H. 2008. A CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications. Palisades, NY: Center for International Earth Science Information Network of Columbia University.

Doll, C. N. H. 2010. Population Detection Profiles of DMSP - OLS night - time imagery by regions of the world. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Doll, C. N. H., Morley, J. G. & Muller, J. P. 2004. Geographic Information Issues associated with Socio-Economic Modelling from Night-time light Remote Sensing Data. *ISPRS Conference.* Istanbul.

Doll, C. N. H., Muller, J.-P. & Elvidge, C. D. 2000. Night-time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions. *AMBIO: A Journal of the Human Environment,* 29**,** 157-162.

Doll, C. N. H. & Muller, J. P. 1999. An evaluation of global urban growth via comparison of DCW and DMSP-OLS satellite data. *In:* International Geoscience and Remote Sensing Symposium (IGARSS) 1134-1136.

Doll, C. N. H., Muller, J. P. & Elvidge, C. D. 2002. Night-time Imagery as a Tool for Global Mapping of Socio-economic parameters and Greenhouse Gas Emissions. *Ambio,* 29**,** 157 - 162.

Doll, C. N. H., Muller, J. P. & Morley, J. G. 2006. Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, 57**,** 75-92.

Drummond, J. 1995. Positional accuracy. *In:* Guptill, S. C. & Morrison, J. L. (eds.) *Elements of Spatial Data Quality*. Oxford, UK: Elsevier Science, published on behalf of the International Cartographic Association.

Duke-Williams, O. & Rees, P. 1998. Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. *International Journal of Geographical Information Science,* 12**,** 579 - 605.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M. & Rosenberg, M. S. 2002. A balanced view of scale in spatial statistical analysis. *Ecography,* 25**,** 626-640.

Ebener, S., Murray, C., Tandon, A. & Elvidge, C. D. 2005. From wealth to health: modeling the distribution of income per capital at the sub-national level using nighttime light imager. *International Journal of Health Geographics,* 4**,** 5.

Elmore, A. J., Mustard, J. F., Manning, S. J. & Lobell, D. B. 2000. Quantifying Vegetation Change in Semiarid Environments: Precision and Accuracy of Spectral Mixture Analysis and the Normalized Difference Vegetation Index. *Remote Sensing of Environment,* 73**,** 87-102.

Elvidge, C., Baugh, K., Tuttle, B. T., Howard, A. T., Pack, D., Milesi, C. & Erwin, E. H. 2007a. A Twelve Year Record of National and Global Gas Flaring Volumes Estimated Using Satellite Data.

Elvidge, C., Hobson, V. R., Nelson, I., Safran, J., Tuttle, B. T., Dietz, J. B. & Baugh, K. 2003. Overview of DMSP OLS and scope of applications. *In:* Mesev, V. (ed.) *Remotely Sensed Cities.* London: Taylor and Francis.

Elvidge, C., Milesi, C., Nemani, R., Tuttle, B. T., Dietz, J. B., Sutton, P. C. & Vogelmann, J. E. 2004. US constructed area approaches the size of Ohio. *EOS Transactions,* 85.

Elvidge, C. D. 1999. AVIRIS observations of nocturnal lighting. *In:* AVIRIS Airborne Geosciences Workshop Proceedings.

Elvidge, C. D. 2010. *RE: Interested in the seminar at Hanoi.* Type to Roychowdhury, K.

Elvidge, C. D., Baugh, K., Safran, J., Tuttle, B. T., Howard, A. T., Hayes, P. J. & Erwin, E. H. 2005. Preliminary Results From Nighttime Lights Change Detection. *In:* 5th International Symposium Remote Sensing of Urban Areas (URS 2005), Tempe, AZ, USA.

Elvidge, C. D., Baugh, K. E., Dietz, J. B., Bland, T., Sutton, P. C. & Kroehl, H. W. 1999. Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements-a new device for portraying the Earth's surface entire. *Remote Sensing of Environment,* 68**,** 77-88.

Elvidge, C. D., Baugh, K. E., Hobson, V. R., Kihn, E. A., Kroehl, H. W., Davis, E. R. & Cocero, D. 1997a. Satellite inventory of human settlements using nocturnal radiation emissions: a contribution for the global toolchest. *Global Change Biology,* 3**,** 387-395.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W. & Davis, E. R. 1997b. Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogrammetric Engineering & Remote Sensing,* 63**,** 727-734.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R. & Davis, C. W. 1997c. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing,* 18**,** 1373 - 1379.

Elvidge, C. D., Cinzano, P., Pettit, D. R., Arvesen, J., Sutton, P., Small, C., Nemani, R., Longcore, T., Rich, C., Safran, J., Weeks, J. & Ebener, S. 2007b. The Nightsat mission concept. *International Journal of Remote Sensing,* 28**,** 2645 - 2670.

Elvidge, C. D., Hobson, V. R., Baugh, K. E., Dietz, J. B., Shimabukuro, Y. E., Krug, T., Novo, E. M. L. M. & Echavarria, F. R. 2001a. DMSP-OLS estimation of tropical forest area impacted by surface fires in Roraima, Brazil: 1995 versus 1998. *International Journal of Remote Sensing,* 22**,** 2661-2673.

Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I. L., Safran, J., Dietz, J. B. & Tuttle, B. T. 2001b. Night-time lights of the world: 1994-1995. *ISPRS Journal of Photogrammetry and Remote Sensing,* 56**,** 81-99.

Elvidge, C. D., Pettit, D., Imhoff, M., Nemani, R., Pack, D. & Cinzano, P. 2008. Observational considerations for moderate resolution nighttime lights. *In:* Earth Observing Systems XIII, San Diego, CA, USA. SPIE, 70810V-9.

Elvidge, C. D., Safran, J., Tuttle, B. T., Sutton, P., Cinzano, P., Pettit, D., Arvesen, J. & Small, C. 2007c. Potential for global mapping of development via a nightsat mission. *GeoJournal,* 69**,** 45.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Baugh, K. E. & Tuttle, B. T. 2009a. Global Urban Mapping Based on Nighttime Lights. *In:* Gamba, P. & Herold, M. (eds.) *Global Mapping of Human Settlement.*

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B. & Bright, E. 2009b. A global poverty map derived from satellite data. *Computers & Geosciences,* 35**,** 1652-1660.

Elvidge, C. D., Sutton, P. C., Pettit, D. R., Cinzano, P. & Small, C. 2007d. Overview of the Nightsat mission concept. *Urban Remote Sensing Joint Event, 2007***,** 1-5.

Elvidge, C. D., Tuttle, B. T., Sutton, P. C., Baugh, K. E., Howard, A. T., Milesi, C., Bhaduri, B. & Nemani, R. 2007e. Global distribution and density of constructed impervious surfaces. *Sensors,* 7**,** 1962-1979.

Encyclopaedia Britannica 2007. *Census,* Chicago, Encyclopaedia Britannica Inc.

Field, A. P. 2009. *Discovering statistics using SPSS,* SAGE publications Ltd.

Fotheringham, A. S. 1989. Scale - independent spatial analysis. *In:* Goodchild, M. F. & Gopal, S. (eds.) *Accuracy of Spatial Databases.*

Fotheringham, A. S. & Wong, D. W. S. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A,* 23**,** 1025-1044.

Frederik, B. 2002. An overview and evaluation of composite indices of development. *Social Indicators Research,* 59**,** 115.

Geoghegan, J., Pritchard Jr, L., Ogneva-Himmelberger, Y., Chowdhury, R. R., Sanderson, S. & Turner Ii, B. L. 1998. 'Socializing The Pixel'and 'Pixelizing The Social'in Land-Use and Land-Cover Change. *People and pixels: Linking remote sensing and social science.* Washington, DC: National Academy Press.

Ghosh, J. K. & Dubey, A. 2008. India's new map policy – Utility of civil users. *Current Science,* 94**,** 332-337.

Ghosh, T., Elvidge, C. D., Sutton, P. C., Baugh, K. & Ziskin, D. 2010. Estimating the Information and Technology Development Index (IDI) Using Nighttime Satellite Imagery. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Goodchild, M. 2009. Epilogue: Putting Research into Practice. *Quality Aspects in Spatial Data Mining.* CRC Press.

Goodchild, M. F. 1995. Attribute accuracy. *In:* Guptill, S. C. & Morrison, J. L. (eds.) *Elements of Spatial Data Quality.* Oxford, UK: Elsevier Science, published on behalf of the International Cartographic Association.

Guptill, S. C. & Morrison, J. L. 1995. *Elements of Spatial Data Quality,* Oxford, UK, Elsevier Science, published on behalf of the International Cartographic Association.

Harvey, J. T. 2002. Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing,* 23**,** 2071-2095.

Healey, J. F. 2008. *Statistics: a tool for social research*, Wadsworth Pub Co.

Henderson, F. M. & Utano, J. J. 1975. Assessing general urban socio-economic conditions with conventional air photography. *Photogrammetria,* 31**,** 81-89.

Henderson, M., Yeh, E. T., Gong, P., Elvidge, C. & Baugh, K. 2003. Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing,* 24**,** 595-609.

Higgs, E. 2004. *The information state in England: The central collection of information on citizens since 1500*, Palgrave Macmillan.

High Energy Astrophysics Archive Research Centre. 2003. *Defense Meteorological Satellite Program (DMSP) Series* [Online]. Available: http://heasarc.nasa.gov/docs/heasarc/missions/dmsp.html [Accessed 8th Sep 2009].

Hsu, F. C., Elvidge, C. D. & Matsuno, Y. 2010. Steel Stock Estimation for Building and Civil Construction by Satellite Image. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Hsu, S. Y. 1973. "Population Estimation from ERTS Imagery: Methodology and Evaluation". *In:* Proceedings of the American Society of Photogrammetry 39th Annual Meeting, 11-16th March Washington D.C., 583-591.

Hunter, G. J. & Goodchild, M. F. 1995. Modeling the Uncertainty of Vector Data in Geographic Information Systems. *In:* ESRI International User Conference,, Palm Springs, CA.

Hunter, G. J., Jones, S. D., Bregt, A. K. & Masters, E. G. 2003a. Spatial Data Quality. *In:* Medeiros, C. M. B. (ed.) *Advanced Geographic Information Systems. Encyclopaedia of Life Support Systems (EOLSS).* Oxford, U.K: Developed under the auspices of the UNESCO. EOLSS Publishers.

Hunter, G. J., Wachowicz, M. & Bregt, A. K. 2003b. Understanding spatial data usability. *Data Science Journal (Spatial Data Usability Special Section),* 2**,** 79-89.

Iisaka, J. & Hegedus, E. 1982. Population estimation from Landsat imagery. *Remote Sensing of Environment,* 12**,** 259-272.

Image Science and Analysis Laboratory, N.-J. S. C., . 2009. *The Gateway to Astronaut Photography of Earth* [Online]. Available: http://eol.jsc.nasa.gov/ [Accessed 9th November 2009].

Imhoff, M. L., Lawrence, W. T., Elvidge, C. D., Paul, T., Levine, E., Privalsky, M. V. & Brown, V. 1997a. Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States. *Remote Sensing of Environment,* 59**,** 105-117.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C. & Elvidge, C. D. 1997b. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. *Remote Sensing of Environment,* 61**,** 361-370.

India Planning Commission. 2007. *Maharashtra, development report,* New Delhi, Academic Foundation under arrangement with Planning Commission, Govt. of India.

Indicus Analytics Private Limited 2009. District Gross Domestic Product series 2001 - 2007.

Innes, P. 2009. *Testing and Fixing for Normality* [Online]. Available: http://www.business.uq.edu.au/download/attachments/6553737/statssheet15.pdf [Accessed 17th July 2010].

Kiran Chand, T. R., Badarinath, K. V. S., Krishna Prasad, V., Murthy, M. S. R., Elvidge, C. D. & Tuttle, B. T. 2006. Monitoring forest fires over the Indian region using Defense Meteorological Satellite Program-Operational Linescan System nighttime satellite data. *Remote Sensing of Environment,* 103**,** 165-178.

Kohiyama, M., Hayashi, H., Maki, N., Higashida, M., Kroehl, H. W., Elvidge, C. D. & Hobson, V. R. 2004. Early damaged area estimation system using DMSP-OLS night-time imagery. *International Journal of Remote Sensing*, 25**,** 2015-2036.

Kruse, F. A. & Elvidge, C. D. 2011. Characterizing urban light sources using imaging spectrometry. *In:* Urban Remote Sensing Event (JURSE), 2011 Joint, 11-13 April 2011. 149-152.

Langford, M., Maguire, D. J. & Unwin, D. J. 1991. The areal interpolation problem: estimating population using remote sensing in a GIS framework. *Handling Geographical Information: Methodology and Potential Applications***,** 55-77.

Lee, T. E., Miller, S. D., Turk, F. J., Schueler, C., Julian, R., Deyo, S., Dills, P. & Wang, S. 2006. The NPOESS VIIRS Day/Night Visible Sensor. *Bulletin of the American Meteorological Society,* 87**,** 191-199.

Li, G. & Weng, Q. 2005. Using Landsat ETM+ imagery to measure population density in Indianapolis, Indiana, USA. *Photogrammetric Engineering and Remote Sensing,* 71**,** 947-958.

Li, G. & Weng, Q. 2007. Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. *International Journal of Remote Sensing,* 28**,** 249-267.

Lieske, R. W. 1981. DMSP primary sensor data acquisition. *In:* International Telemetering conference. 1013-1020.

Lillesand, T. M., Kiefer, R. W. & Chipman, J. W. 2004. *Remote sensing and image interpretation,* New York ; [Great Britain], Wiley.

Liu, X. 2003. Estimation of the spatial distribution of urban population using high spatial resolution satellite imagery. *Estimation of the Spatial Distribution of Urban Population Using High Spatial Resolution Satellite Imagery***,** 175.

Lo, C. 2008a. Population Estimation Using Geographically Weighted Regression. *GIScience & Remote Sensing,* 45**,** 131-148.

Lo, C. P. 1979. Surveys of squatter settlements with sequential aerial photography - A case study in Hong Kong. *Photogrammetria,* 35**,** 45-63.

Lo, C. P. 1995. Automated population and dwelling unit estimation from high- resolution satellite images: a GIS approach. *International Journal of Remote Sensing,* 16**,** 17-34.

Lo, C. P. 1997. Application of Landsat TM data for quality of life assessment in an urban environment. *Computers, Environment and Urban Systems,* 21**,** 259-276.

Lo, C. P. 2001a. *Estimating the Population of China using DMSP Operational Linescan System Nighttime Images* [Online]. Available: http://www.gis.usu.edu/docs/protected/procs/asprs/asprs2001/Proceed/00062.pdf [Accessed 2nd September 2009].

Lo, C. P. 2001b. Modeling the population of China using DMSP operational linescan system nighttime data. *Photogrammetric Engineering and Remote Sensing,* 67**,** 1037-1047.

Lo, C. P. 2002. Urban Indicators of China from Radiance-Calibrated Digital DMSP-OLS Nighttime Images. *Annals of the Association of American Geographers,* 92**,** 225-240.

Lo, C. P. 2006. Estimating Population and Census Data. *In:* Ridd, M. K. & Hipple, J. D. (eds.) *Remote Sensing of Human Settlements.* 3 ed. Bethesda, Maryland: American Society for Photogrammetry and Remote Sensing.

Lo, C. P. 2008b. Population estimation using geographically weighted regression. *GIScience and Remote Sensing,* 45**,** 131-148.

Lo, C. P. & Faber, B. J. 1997. Integration of landsat thematic mapper and census data for quality of life assessment. *Remote Sensing of Environment,* 62**,** 143-157.

Longcore, T. & Rich, C. 2004. Ecological light pollution. *Frontiers in Ecology and the Environment,* 2**,** 191-198.

Lu, D., Weng, Q. & Li, G. 2006. Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing,* 27**,** 3553 - 3570.

Lu, Z., Im, J., Quackenbush, L. & Halligan, K. 2010. Population estimation based on multi-sensor data fusion. *International Journal of Remote Sensing,* 31**,** 5587 - 5604.

Maheshwari, S. 1996. *The census administration under the raj and after,* Concept Pub. Co.

Marceau, D. J. 1999. The scale issue in social and natural sciences. *Canadian Journal of Remote Sensing,* 25**,** 347-356.

Marceau, D. J. & Hay, G. J. 1999. Remote Sensing Contributions to the Scale Issue. *Canadian Journal of Remote Sensing,* 25**,** 357-366.

Martin, D. 1989. Mapping Population Data from Zone Centroid Locations. *Transactions of the Institute of British Geographers,* 14**,** 90 - 97.

Martin, D. 1992. Postcodes and the 1991 Census of Population: Issues, Problems and Prospects. *Transactions of the Institute of British Geographers,* 17**,** 350-357.

Martin, D. 2000. Towards the Geographies of the 2001 UK Census of Population. *Transactions of the Institute of British Geographers,* 25**,** 321-332.

Martin, D. 2006. Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers,* 31**,** 6-18.

Martin, D. & Bracken, I. 1993. The integration of socioeconomic and physical resource data for applied land management information systems. *Applied Geography,* 13**,** 45-53.

Massey, J. 2002. Automated interpretation of the spatial distribution of socio-economic conditions *GIS Development.*

McCloy, K. R. 2006. *Resource Management Information Systems: Remote Sensing, GIS and Modelling,* Boca Raton, CRC Press.

McCoy, R. M. 2005. *Field Methods in Remote Sensing,* New York, The Guilford Press.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B. L., M. A., Cohen, W. B., Harmon, M. E. & McKee, W. A. 1999. Use of Large-Footprint Scanning Airborne Lidar To Estimate Forest Stand Characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment,* 67**,** 298-308.

Mesev, V. 1998. The use of census data in urban image classification. *Photogrammetric Engineering and Remote Sensing,* 64**,** 431-436.

Metivier, E. D. & McCoy, R. M. 1971. Mapping urban poverty housing from aerial photographs. *Proceedings of the Seventh International Symposium on Remote Sensing of Environment***,** 1563-1569.

Milesi, C., Elvidge, C. D., Nemani, R. & Running, S. W. 2003. Assessing the impact of urban land development on net primary productivity in the southeastern United States. *Remote Sensing of Environment,* 86**,** 401-410.

Miller, S. D., Lee, T. F., Turk, F. J., Kuciauskas, A. P. & Hawkins, J. D. 2005. Shedding new light on nocturnal monitoring of the environment with the VIIRS day/night band. *In:* Proceedings of SPIE - The International Society for Optical Engineering. 1-9.

Miller, S. D. & Turner, R. E. 2009. A dynamic lunar spectral irradiance data set for NPOESS/VIIRS day/night band night time environmental applications. *IEEE Transactions on Geoscience and Remote Sensing,* 47**,** 2316-2329.

Montevecchi, W. 2006. Influences of artificial light on marine birds. *Ecological Consequences of Artificial Night Lighting***,** 94-113.

Morgan, R. W. 1984. Further Comments on "the census". *Photogrammetric Engineering and Remote Sensing,* 50**,** 80.

Morrison, J. L. 1995. Spatial data quality. *In:* Guptill, S. C. & Morrison, J. L. (eds.) *Elements of Spatial Data Quality.* Oxford, UK: Elsevier Science, published on behalf of the International Cartographic Association.

Morrow-Jones, H. A. & Watkins, J. F. 1984. Remote Sensing Technology and the US Census. *Photogrammetric Engineering and Remote Sensing,* 50**,** 229-232.

Morton, T. A. & Yuan, F. 2009. Analysis of population dynamics using satellite remote sensing and US census data. *Geocarto International,* 24**,** 143-163.

Nagatani, I. 2010. A methodology to create DMSP-OLS night-time mosaic image for monitoring fishing boats. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

NASA. 2009. *Cities at night: The view from Space* [Online]. Available: http://earthobservatory.nasa.gov/Features/CitiesAtNight/ [Accessed 9th November 2009].

National Geophysical Data Centre. 2006. *DMSP data Download* [Online]. Available: http://www.ngdc.noaa.gov/dmsp/download_rad_cal_96-97.html [Accessed 25th May 2008].

National Geophysical Data Centre. 2007a. *Defense Meteorological Satellite Program (DMSP) Data Archive, Research, and Products* [Online]. Available: http://www.ngdc.noaa.gov/dmsp/dmsp.html [Accessed 20th April].

National Geophysical Data Centre. 2007b. *DMSP Archive Description* [Online]. Available: http://www.ngdc.noaa.gov/dmsp/archive.html [Accessed 20th April 2008].

National Geophysical Data Centre. 2007c. *OLS - Operational Linescan System* [Online]. Available: http://www.ngdc.noaa.gov/dmsp/sensors/ols.html [Accessed 20th April 2008].

Nghiem, S. V., Balk, D., Rodriguez, E., Neumann, G., Sorichetta, A., Small, C. & Elvidge, C. D. 2009. Observations of urban and suburban environments with global satellite scatterometer data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64**,** 367-380.

Nordbeck, S. 1965. The law of allometric growth. *Michigan Inter-University Community of Mathematical Geographers Paper*, 7.

Office for National Statistics. 2008. *Census History* [Online]. Available: http://www.ons.gov.uk/census/census-history/index.html [Accessed 29th November 2009].

Office of the Registrar General and Census Commisioner of India. 2001a. *About Census* [Online]. Available: www.censusindia.gov.in/Census_and_You/about_census.aspx [Accessed 11th September, 2008].

Office of the Registrar General and Census Commisioner of India. 2001b. *History of Indian census* [Online]. Available: http://censusindia.gov.in/Data_Products/Library/Indian_perceptive_link/History_link/censushistory.htm [Accessed 29th November 2009].

Office of the Registrar General and Census Commissioner 2011a. Census 2011: Provisional Population Totals.

Office of the Registrar General and Census Commissioner. 2011b. Organizing Census. *http://censusindia.gov.in/Ad_Campaign/drop_in_articles/01-Organizing_Census.pdf* [Online]. [Accessed 3rd June, 2011].

Office of the Registrar General and Census Commissioner. 2011c. The Relevance and Ramifications of Census. *http://censusindia.gov.in/Ad_Campaign/drop_in_articles/02-Relevance_and_ramifications_of_Census.pdf* [Online]. [Accessed 3rd June, 2011].

Openshaw, S. 1984. The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38**,** 41.

Paul, C. K. 1984. More on the census. *Photogrammetric Engineering and Remote Sensing*, 50**,** 1610 - 1612.

Perry, G., Buchanan, B. W., Fisher, R. N., Salmon, M. & Sharon, E. W. 2008. Effects of Artificial Night Lighting on Amphibians and Reptiles in Urban Environments. *In:* Mitchell, J. C., Jung Brown, R. E. & Bartholomew, B. (eds.) *Urban Herpetology*. Society for the Study of Amphibians and Reptiles.

Porter, W. M. & Enmark, H. T. 1987. A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS). *Proc. SPIE,* 834**,** 22-31.

Potere, D. & Schneider, A. 2007. A critical look at representations of urban areas in global maps. *GeoJournal*, 69**,** 55-80.

Pozzi, F., Small, C. & Yetman, G. 2003. Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world. *Earth Observation Magazine,* 12**,** 24-30.

Qi, X. & Chopping, M. 2007. Expansion of Urban Area in the Yellow River Zone, Inner Mongolia Autonomous Region, China, from DMSP OLS Nighttime Lights Data. *IEEE International Geoscience & Remote Sensing Symposium.* Barcelona, Spain.

Reinke, K. & Jones, S. 2006. Integrating vegetation field surveys with remotely sensed data. *Ecological Management & Restoration*, 7**,** S18-S23.

Robinson, W. S. 1950. Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15**,** 351-357.

Roy Chowdhury, P. K. & Maithani, S. 2010. Monitoring growth of built-up areas in indo-gangetic plain using multi-sensor remote sensing data. *Journal of the Indian Society of Remote Sensing,* 38**,** 291-300.

Roychowdhury, K., Jones, S. & Arrowsmith, C. 2009. Assessing the utility of DMSP/OLS night-time images for characterizing indian urbanization. *In:* Urban Remote Sensing Event, 2009 Joint. 1-7.

Roychowdhury, K., Jones, S., Arrowsmith, C. & Reinke, K. 2011a. Indian census using satellite images: Can DMSP-OLS data be used for small administrative regions? . *In:* Urban Remote Sensing Event, 2011 Joint, Munich.

Roychowdhury, K., Jones, S. D., Arrowsmith, C. & Reinke, K. 2011b. A Comparison of High and Low Gain DMSP/OLS Satellite Images for the Study of Socio-Economic Metrics. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* 4**,** 35 - 42.

Roychowdhury, K., Jones, S. D., Arrowsmith, C., Reinke, K. & Bedford, A. 2010. Estimating census metrics at a sub-national level using radiance calibrated DMSP-OLS night-time images. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Roychowdhury, K., Taubenboeck, H. & Jones, S. D. 2011c. Delineating urban, suburban and rural areas using Landsat and DMSP-OLS night-time images. *In:* Urban Remote Sensing Event (JURSE), 2011 Joint, 11-13 April 2011. 33-36.

Salmon, M. 2005. Protecting sea turtles from artificial night lighting at Florida's oceanic beaches. *Ecological Consequences of Artificial Night Lighting***,** 141-168.

Salmon, M., Witherington, B. E. & Elvidge, C. D. 2000. Artificial lighting and the recovery of sea turtles. *Sea Turtles of the Indo-Pacific: Research, Management and Conservation*.

Saxon, E. C., Parris, T. & Elvidge, C. D. 1997. Satellite Surveillance of National CO2 Emissions from Fossil Fuels. Harvard Institute for International Development, Harvard University, USA.

Schneider, A., Friedl, M. A., McLver, D. M. & Woodcock, C. E. 2003. Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data. *Photogrammetric Engineering and Remote Sensing,* 69**,** 1377-1386.

Schneider, A. & Woodcock, C. E. 2008. Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies,* 45**,** 659-692.

Schueler, C. F., Clement, J. E., Miller, S. W., Kealy, P., Ardanuy, P., Cota, S. A., DeLuccia, F., Haas, J. M., Mango, S., Speidel, C. K. & Swenson, H. 2002. NPOESS VIIRS: Next generation polar-orbiting atmospheric imager. *In:* Proceedings of SPIE - The International Society for Optical Engineering. 50-64.

Scott, J. C. 1999. *Seeing like a state: How certain schemes to improve the human condition have failed,* Yale Univ Pr.

Shamasastry, R. 1915. Kautilya Arthashastra. Oriental Research Institute of Mysore.

Small, C., Pozzi, F. & Elvidge, C. D. 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment,* 96**,** 277-291.

Stewart, J. Q. & Warntz, W. 1958. Physics of population distribution. *Journal of regional science,* 1**,** 99-121.

Sullivan, W. T. 1989. A 10 km resolution image of the entire night-time Earth based on cloud-free satellite photographs in the 400 - 1100 nm band. *International Journal of Remote Sensing,* 10**,** 1 - 5.

Survey of India. 2005. *National Map Policy* [Online]. Available: http://www.surveyofindia.gov.in/tenders/National%20Map%20Policy.htm [Accessed 23rd November 2010].

Sutton, P. 2003a. Estimation of human population parameters using night-time satellite imagery. *In:* Mesev, V. (ed.) *Remotely Sensed Cities.* London, UK: Taylor and Francis.

Sutton, P., Roberts, D., Elvidge, C. & Meij, H. 1997. A comparison of nighttime satellite imagery and population density for the continental United States. *Photogrammetric Engineering and Remote Sensing,* 63**,** 1303-1313.

Sutton, P., Roberts, D., Elvidge, C. D. & Baugh, K. 2001. Census from Heaven: an estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing,* 22**,** 3061-3076.

Sutton, P. C. 1997. Modeling population density with night-time satellite imagery and GIS. *Computers, Environment and Urban Systems,* 21**,** 227-244.

Sutton, P. C. 2003b. A scale-adjusted measure of "Urban sprawl" using nighttime satellite imagery. *Remote Sensing of Environment,* 86**,** 353-369.

Sutton, P. C. & Costanza, R. 2002. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics,* 41**,** 509-527.

Sutton, P. C., Cova, T. J. & Elvidge, C. D. 2006. Mapping “Exurbia” in the Conterminous United States Using Nighttime Satellite Imagery. *Geocarto International,* 21.

Sutton, P. C., Elvidge, C. D., Tuttle, B. T., Ziskin, D., Baugh, K. & Ghosh, T. 2010. A 2010 Mapping of the Constructed Surface Area Density for S.E. Asia - Preliminary Results. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

Thomlinson, J. R., Bolstad, P. V. & Cohen, W. B. 1999. Coordinating Methodologies for Scaling Landcover Classification from Site-Specific to Global: Steps toward Validating Global Map Products. *Remote Sensing of Environment,* 70**,** 16-28.

Tobler, W. R. 1969. Satellite confirmation of settlement size coefficients. *Area,* 1**,** 30-34.

Toenges-Schuller, N., Stein, O., Rohrer, F., Wahner, A., Richter, A., Burrows, J. P., Beirle, S., Wagner, T., Platt, U. & Elvidge, C. D. 2006. Global distribution pattern of anthropogenic nitrogen oxide emissions: Correlation analysis of satellite measurements and model calculations. *Journal of Geophysical Research,* 111.

Truong, K. Y. 2007. *Some Comments on Regression Without Intercept* [Online]. The University of North Carolina at Chapel Hill. Available: http://www.bios.unc.edu/~truong/b663/pdf/noint.pdf [Accessed 26th October 2010].

Turner, M. G. 1990. Spatial and temporal analysis of landscape patterns. *Landscape Ecology,* 4**,** 21-30.

Tuttle, B. T. May 2008. *RE: Personal communication.* Type to Roychowdhury, K.

Tuttle, B. T. August 2009. *RE: Confusion with data processing.* Type to Roychowdhury, K.

Tuxbury, S. M. & Salmon, M. 2005. Competitive interactions between artificial lighting and natural cues during seafinding by hatchling marine turtles. *Biological Conservation,* 121**,** 311-316.

United Nations Human Settlements Programme. 2006. *The state of the world's cities 2006/2007 the millennium development goals and urban sustainability 30 years of shaping the Habitat Agenda,* London Earthscan.

Unwin, D. J. 1996. GIS, spatial analysis and spatial statistics. *Progress in Human Geography,* 20**,** 540.

Visvalingam, M. 1991. Areal units and the linking of data: some conceptual issues. *In:* Worrall, L. (ed.) *Spatial Analysis and Spatial Policy using Geographic Information Systems.* Belhaven Press.

Vogelman, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K. & Van Driel, N. 2001. Completion of the 1990 s National Land Cover Data set for the conterminous United States for

Landsat Thematic Mapper data and ancillary data sources. *PE & RS- Photogrammetric Engineering and Remote Sensing,* 67**,** 650-655.

Walsh, S. J., Evans, T. P., Welsh, W. F., Entwisle, B. & Rindfuss, R. R. 1999. Scale-dependent relationships between population and environment in northeastern Thailand. *Photogrammetric Engineering and Remote Sensing,* 65**,** 97-105.

Waluda, C. M., Yamashiro, C., Elvidge, C. D., Hobson, V. R. & Rodhouse, P. G. 2004. Quantifying light-fishing for Dosidicus gigas in the eastern Pacific using satellite remote sensing. *Remote Sensing of Environment,* 91**,** 129-133.

Weber, C. & Hirsch, J. 1992. Some urban measurements from SPOT data: urban life quality indices. *International Journal of Remote Sensing,* 13**,** 3251-3261.

Weber, K. T., Théau, J. & Serr, K. 2008. Effect of coregistration error on patchy target detection using high-resolution imagery. *Remote Sensing of Environment,* 112**,** 845-850.

Webster, C. J. 1996. The potential of urban texture measures in monitoring urbanisation from space. *In:* Fung, T., Lai, P. C., H, L. & Yeh, A. G. O. (eds.) *GIS in Asia.* Singapore: GIS Asia Pacific.

Welch, R. 1980. Monitoring urban population and energy utilization patterns from satellite data. *Remote Sensing of Environment,* 9**,** 1-9.

Welch, R. & Zupko, S. 1980. Urbanized area energy utilization patterns from DMSP data. *Photogrammetric Engineering and Remote Sensing,* 46**,** 201-207.

Wiese, F. K., Montevecchi, W. A., Davoren, G. K., Huettmann, F., Diamond, A. W. & Linke, J. 2001. Seabirds at risk around offshore oil platforms in the North-west Atlantic. *Marine Pollution Bulletin,* 42**,** 1285-1290.

Wrigley, N., Holt, T., Steel, D. & Tranmer, M. 1996. Analysing, modelling, and resolving the ecological fallacy. *Spatial Analysis, Modelling in a GIS Environment.* GeoInformation International.

Wu, S. S., Wang, L. & Qiu, X. 2008. Incorporating GIS building data and census housing statistics for sub-block-level population estimation. *Professional Geographer,* 60**,** 121-135.

Wunsch, D. R. & Gades, R. E. 1986. Survey Research: Determining Sample Size and Representative Response. and The Effects of Computer Use on Keyboarding Technique and Skill. 40**,** 31-36.

Yin, Z. Y., Stewart, D. J., Bullard, S. & MacLachlan, J. T. 2005. Changes in urban built-up surface and population distribution patterns during 1986-1999: A case study of Cairo, Egypt. *Computers, Environment and Urban Systems,* 29**,** 595-616.

Yuan, Y., Smith, R. M. & Limp, W. F. 1997. Remodeling census population with spatial information from Landsat TM imagery. *Computers, Environment and Urban Systems,* 21**,** 245-258.

Zar, J. H. 2010. *Biostatistical Analysis,* Upper Saddle River, New Jersy, Pearson Education Inc.

Zhuo, L., Ichinose, T., Zheng, J., Chen, J., Shi, P. J. & Li, X. 2009. Modelling the population density of China at the pixel level based on DMSP/OLS non-radiance-calibrated night-time light images. *International Journal of Remote Sensing,* 30**,** 1003 - 1018.

Ziskin, D., Baugh, K., Hsu, F. C., Ghosh, T. & Elvidge, C. D. 2010. Methods used for the 2006 Radiance Lights. *In:* Elvidge, C., ed. Asia-Pacific advanced Network (APAN) Workshop, Hanoi, Vietnam.

10. Appendix

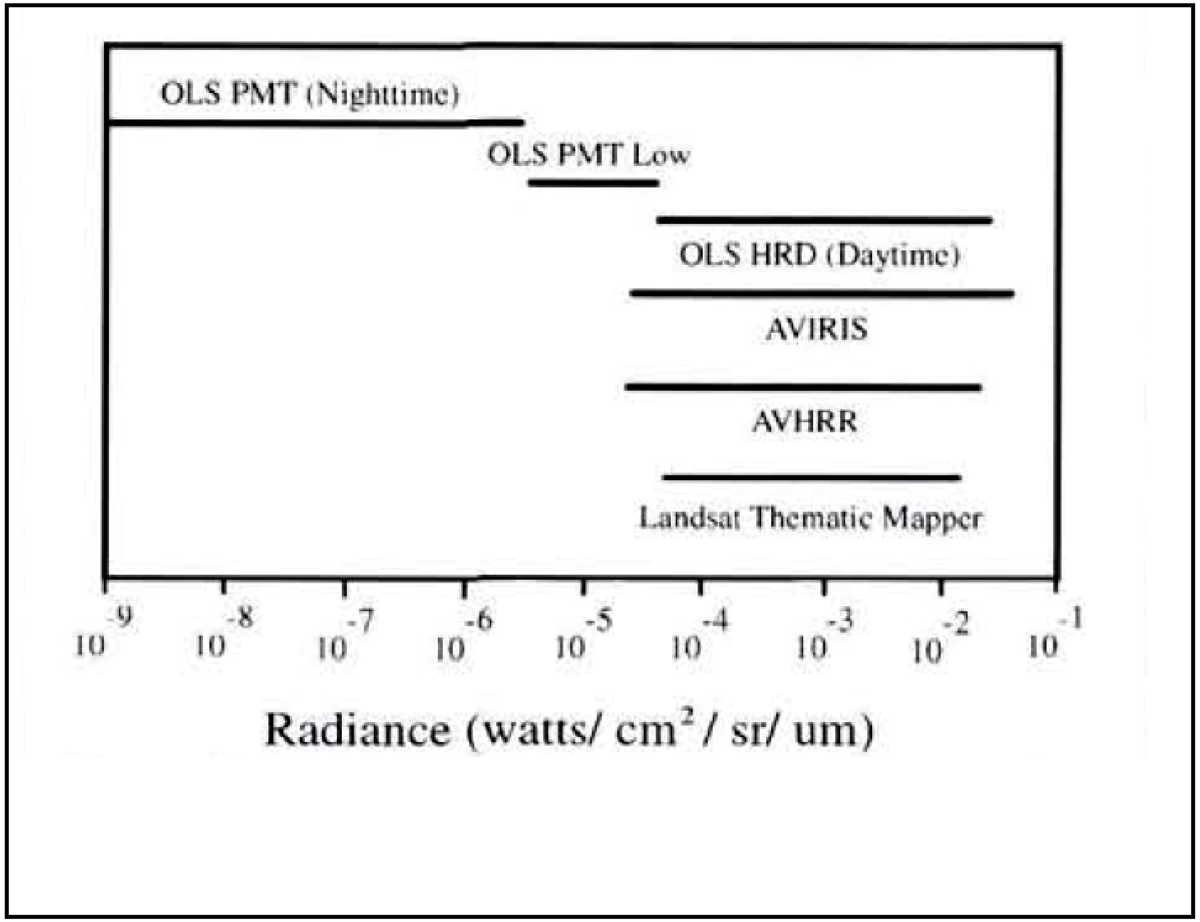


Figure 10. 1: Comparison of VNIR Dynamic ranges of DMSP/OLS, NOAA-AVHRR, NRSA-AVIRIS and Landsat Thematic Mapper (Elvidge et al., 1997b p. 728)

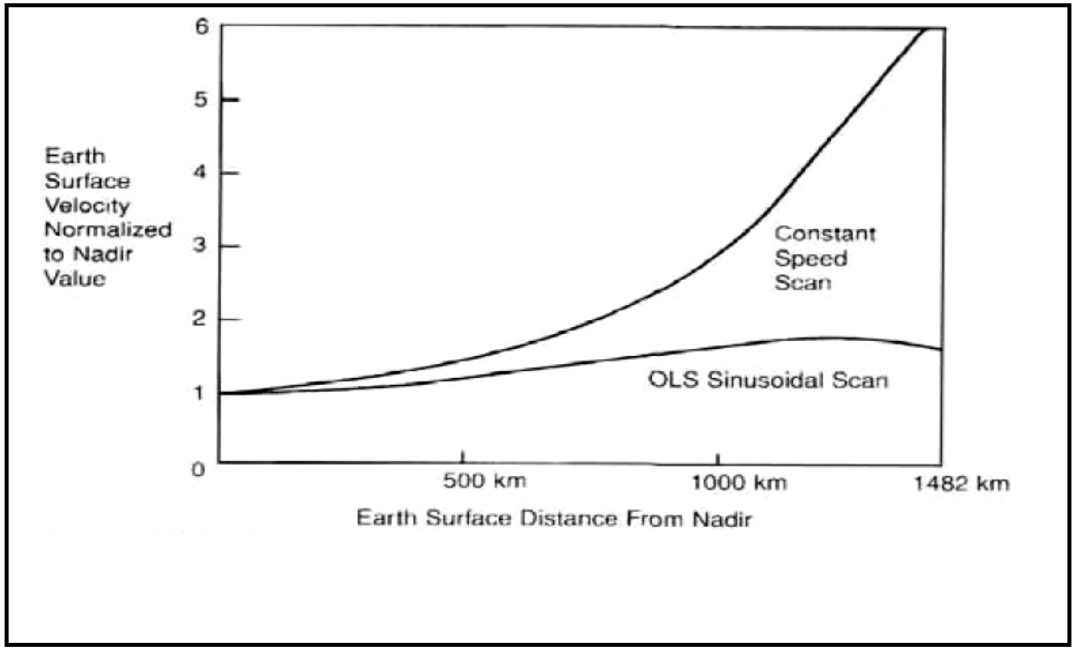


Figure 10. 2: Uniformity of DMSP/OLS Earth surface Scan velocity. Figure courtesy of Westinghouse Electric Corporation (Elvidge et al., 1997b p. 729)

Table 10.1: Summary statistics of all the variables

	Mean	Minimum	Maximum	Lower	Upper	Range	Variance	Std.Dev.	Skewness	Std.Err.	Kurtosis	Std.Err.	Z_Skewness	Z_Kurtosis
Number of Households	460984	180695.0	1517041	247571	560252	1336346	8.832768E+10	297200	2.11234	0.472261	6.03474	0.917777	4.472828	6.575392
Number of Households per square kilometre	51	14.6	97	40	55	82	3.404822E+02	18	1.12614	0.472261	1.80951	0.917777	2.384582	1.971625
Total Population	2290866	868825.0	7232555	1218722	2853004	6363730	2.067464E+12	1437868	1.90870	0.472261	4.98121	0.917777	4.041627	5.427475
Total Population per square kilometre	254	67.3	462	206	287	395	8.022290E+03	90	0.89601	0.472261	1.49679	0.917777	1.89728	1.630881
Total Male	1174887	417890.0	3769128	618912	1456845	3351238	5.663391E+11	752555	1.92736	0.472261	5.07112	0.917777	4.08113	5.525439
Total Male per square kilometre	130	34.1	241	105	149	207	2.215188E+03	47	0.91309	0.472261	1.41112	0.917777	1.933452	1.537546
Total Female	1115979	450935.0	3463427	599810	1396159	3012492	4.703256E+11	685803	1.88302	0.472261	4.86236	0.917777	3.987255	5.297977
Total Female per square kilometre	124	33.2	223	100	138	190	1.818057E+03	43	0.86121	0.472261	1.56278	0.917777	1.823587	1.702784
Total Population (0 - 6 yrs)	328912	105518.0	968851	169145	409643	863333	3.653593E+10	191144	1.71626	0.472261	4.27670	0.917777	3.634126	4.659849
Male (0 - 6 yrs)	172173	54277.0	509367	87072	220462	455090	1.027787E+10	101380	1.68453	0.472261	4.09945	0.917777	3.566944	4.466717
Female (0 - 6 yrs)	156738	51241.0	459484	82073	191734	408243	8.071137E+09	89840	1.75107	0.472261	4.46122	0.917777	3.707848	4.860893
Population (Scheduled Caste)	270216	24515.0	761857	133866	390216	737342	3.988900E+10	199722	1.08356	0.472261	0.64289	0.917777	2.294402	0.700489
Male (Scheduled Caste)	138167	12254.0	390965	68848	201070	378711	1.056803E+10	102801	1.08383	0.472261	0.63098	0.917777	2.294972	0.687504
female (Scheduled Caste)	132048	12261.0	370892	65317	189146	358631	9.395875E+09	96932	1.08303	0.472261	0.65462	0.917777	2.293286	0.713262

Population (Scheduled Tribe)	200435	4952.0	859574	39970	364115	854622	4.542766E+10	213138	1.51587	0.472261	2.41701	0.917777	3.209822	2.633546
Male (Scheduled Tribe)	101794	2515.0	427858	20585	184617	425343	1.150101E+10	107243	1.46118	0.472261	2.11970	0.917777	3.094007	2.309607
Female (Scheduled Tribe)	98641	2437.0	431716	19385	179498	429279	1.122079E+10	105928	1.57316	0.472261	2.73118	0.917777	3.33112	2.975867
Literate Population	1482086	490121.0	5039290	789538	1882748	4549169	1.044410E+12	1021964	2.01945	0.472261	5.42809	0.917777	4.276142	5.914392
Literate Populationper square kilometre	163	34.0	322	117	185	288	4.575255E+03	68	1.02508	0.472261	1.31475	0.917777	2.170572	1.432539
Male Literates	856795	296314.0	2879761	449194	1068415	2583447	3.388896E+11	582142	1.99237	0.472261	5.33898	0.917777	4.218793	5.817295
Male Literates per square kilometre	94	20.6	184	71	107	164	1.445364E+03	38	1.00065	0.472261	1.32933	0.917777	2.118847	1.448419
Female Literates	625291	193807.0	2159529	327605	831961	1965722	1.949436E+11	441524	2.03607	0.472261	5.44218	0.917777	4.311329	5.929745
Female Literates per square kilometre	69	13.4	138	49	81	125	8.968403E+02	30	1.01581	0.472261	1.20958	0.917777	2.150942	1.317949
Illiterate Population	808779	255906.0	2193265	462012	1005963	1937359	1.936866E+11	440098	1.45154	0.472261	3.00937	0.917777	3.073588	3.278977
Male Illiterates	318092	89691.0	889367	181124	399802	799676	3.199081E+10	178860	1.50551	0.472261	3.34304	0.917777	3.187872	3.642543
Female Illiterates	490687	166215.0	1303898	280888	600974	1137683	6.860116E+10	261918	1.40450	0.472261	2.73820	0.917777	2.973995	2.983512
Total Workers	996781	404985.0	2954482	564974	1240258	2549497	3.412887E+11	584199	1.78887	0.472261	4.27480	0.917777	3.787895	4.65778
Total Workers per square kilometre	111	34.5	215	91	121	181	1.335974E+03	37	1.09307	0.472261	2.72650	0.917777	2.314544	2.970769
TOT_WORK_M	616142	228357.0	2016053	331737	749813	1787696	1.622256E+11	402772	1.99303	0.472261	5.30863	0.917777	4.220197	5.784227
Total Male Workers per square kilometre	68	18.9	133	54	74	114	6.588689E+02	26	1.18919	0.472261	1.99607	0.917777	2.518081	2.174892

Total Female Workers	380639	176628.0	938429	229282	445369	761801	3.551296E+10	188449	1.33290	0.472261	1.92335	0.917777	2.822384	2.095659
Total Female Workers per square kilometre	43	15.6	82	38	48	66	1.534621E+02	12	0.97083	0.472261	3.79742	0.917777	2.055712	4.137625
Total Main workers	826739	233827.0	2645429	422084	1035260	2411602	2.830812E+11	532054	1.85071	0.472261	4.81988	0.917777	3.918834	5.251692
Total Main workers per square kilometre	91	23.4	182	75	101	158	1.136925E+03	34	1.03786	0.472261	2.40047	0.917777	2.197645	2.615521
MAINWORK_M	554440	161489.0	1883972	273809	681994	1722483	1.445493E+11	380196	1.99919	0.472261	5.49624	0.917777	4.233226	5.988648
Total Male Main workers per square kilometre	61	14.9	123	48	68	108	6.059622E+02	25	1.16635	0.472261	1.97129	0.917777	2.469715	2.147894
Total Female Main workers	272299	72338.0	761457	155171	345488	689119	2.469107E+10	157134	1.42691	0.472261	2.72179	0.917777	3.021451	2.965638
Total Female Main workers per square kilometre	30	8.5	59	26	35	50	1.010033E+02	10	0.62420	0.472261	2.64642	0.917777	1.321724	2.88351
Total Main Casual Labourer	281879	105965.0	709537	156980	370042	603572	2.739766E+10	165522	1.05305	0.472261	0.44205	0.917777	2.229803	0.481655
Male Main Casual Labourer	171852	62843.0	391397	98531	228741	328554	8.593118E+09	92699	0.93388	0.472261	0.00330	0.917777	1.977461	0.003601
Female Main Casual Labourer	110028	29423.0	318140	48971	164755	288717	5.740512E+09	75766	1.13238	0.472261	0.82569	0.917777	2.397783	0.899666
Total Main Agricultural Labourer	228916	23963.0	546762	152026	274712	522799	1.658081E+10	128766	0.77349	0.472261	0.38462	0.917777	1.637838	0.419083

Male Main Agricultural Labourer	118679	15387.0	290286	78207	142637	274899	4.753165E+09	68943	0.92487	0.472261	0.56985	0.917777	1.958388	0.620907
Female Main Agricultural Labourer	110237	8576.0	256476	73819	132742	247900	3.693867E+09	60777	0.63845	0.472261	0.25020	0.917777	1.351901	0.272612
Total Main Household Labourer	19507	4573.0	69419	7819	26101	64846	3.073057E+08	17530	1.74648	0.472261	2.49698	0.917777	3.698135	2.720684
Male Main Household Labourer	11162	3252.0	35752	5661	15029	32500	6.805972E+07	8250	1.66380	0.472261	2.58037	0.917777	3.523047	2.811549
Female Main Household Labourer	8344	1321.0	49604	2333	8660	48283	1.214358E+08	11020	2.74649	0.472261	8.35697	0.917777	5.81563	9.10566
Main Other Workers (Persons)	296437	64269.0	1597113	113044	352832	1532844	1.102887E+11	332097	2.96520	0.472261	10.29717	0.917777	6.278736	11.21969
Main Other Workers (Male)	252747	56143.0	1323461	95305	305424	1267318	7.605637E+10	275783	2.89776	0.472261	9.88157	0.917777	6.135925	10.76685
Main Other Workers (Female)	43690	6455.0	273652	17075	44198	267197	3.225431E+09	56793	3.24050	0.472261	12.03644	0.917777	6.861674	13.11477
Marginal Workers (Persons)	170042	47140.0	309053	122974	230522	261913	4.949718E+09	70354	0.14592	0.472261	-0.60990	0.917777	0.30899	-0.66454
Marginal Workers (Persons)per square kilometre	20	9.8	42	12	25	32	7.931656E+01	9	0.86573	0.472261	0.00501	0.917777	1.833159	0.005463
Marginal Workers (Male)	61702	16532.0	132081	41417	77016	115549	8.479570E+08	29120	0.70952	0.472261	0.56738	0.917777	1.502396	0.618209

Marginal Workers (Male)per square kilometre	7	3.4	16	4	8	13	1.187231E+01	3	1.01698	0.472261	0.37746	0.917777	2.153436	0.411274
Marginal Workers (Female)	108340	30608.0	180547	78546	134811	149939	1.930903E+09	43942	0.10838	0.472261	-0.58384	0.917777	0.229491	-0.63614
Marginal Workers (Female)per square kilometre	13	6.3	26	8	17	19	3.275346E+01	6	0.80211	0.472261	-0.31277	0.917777	1.698452	-0.34079
Marginal Casual Labourer (Persons)	44400	8773.0	115079	24265	58355	106306	1.028106E+09	32064	1.14509	0.472261	0.14141	0.917777	2.424706	0.154082
Marginal Casual Labourer (Male)	13785	3151.0	29827	7001	19727	26676	8.258878E+07	9088	0.79397	0.472261	-0.76164	0.917777	1.68122	-0.82988
Marginal Casual Labourer (Female)	30615	5584.0	85252	15630	39845	79668	5.471567E+08	23391	1.26437	0.472261	0.55033	0.917777	2.677262	0.59963
Marginal Agricultural Labourer (Persons)	90050	28025.0	184266	62297	117177	156241	1.330903E+09	36482	0.49583	0.472261	0.32318	0.917777	1.049899	0.352132
Marginal Agricultural Labourer (Male)	27623	8819.0	58887	17792	34627	50068	1.553770E+08	12465	0.72379	0.472261	0.08189	0.917777	1.532607	0.089222
Marginal Agricultural Labourer (Female)	62427	19206.0	125379	44337	80552	106173	6.115154E+08	24729	0.36792	0.472261	0.29708	0.917777	0.779069	0.323696
Marginal Household Labourer (Persons)	6853	1082.0	18255	2946	9606	17173	2.729766E+07	5225	1.03658	0.472261	-0.03318	0.917777	2.194926	-0.03615
Marginal Household	1762	331.0	4868	811	2769	4537	1.408106E+06	1187	0.90614	0.472261	0.28801	0.917777	1.918725	0.313812

Labourer (Male)														
Marginal Household Labourer (Female)	5091	741.0	15389	1988	6473	14648	1.732029E+07	4162	1.21559	0.472261	0.49651	0.917777	2.573971	0.540995
Marginal Other Worker (Persons)	28739	3856.0	110327	13451	33088	106471	6.898382E+08	26265	2.04307	0.472261	4.02952	0.917777	4.326142	4.390525
Marginal Other Worker (Male)	18532	2779.0	73630	9350	21895	70851	2.998003E+08	17315	2.31319	0.472261	5.38381	0.917777	4.898123	5.866142
Marginal Other Worker (Female)	10207	1077.0	36697	4888	12195	35620	9.291147E+07	9639	1.80118	0.472261	2.51408	0.917777	3.813942	2.739311
Non Workers (Persons)	1294084	463840.0	4278073	653748	1615847	3814233	7.401908E+11	860343	1.98323	0.472261	5.32265	0.917777	4.199433	5.799499
Non Workers (Persons)per square kilometre	142	32.8	273	113	172	241	2.968901E+03	54	0.91238	0.472261	1.23979	0.917777	1.931935	1.350865
Non Workers (Male)	558745	189533.0	1753075	285096	707032	1563542	1.234587E+11	351367	1.83860	0.472261	4.69146	0.917777	3.893182	5.111764
Non Workers (Male)per square kilometre	62	15.2	112	51	73	97	4.724482E+02	22	0.62587	0.472261	0.95676	0.917777	1.32527	1.042472
Non Workers (Female)	735340	254479.0	2524998	361958	924264	2270519	2.605037E+11	510396	2.07499	0.472261	5.70868	0.917777	4.393731	6.220121
Non Workers (Female)per square kilometre	81	17.7	161	63	96	144	1.088142E+03	33	1.08018	0.472261	1.43095	0.917777	2.287262	1.559152
SEXRATIO	962	919.0	1136	935	977	217	2.548254E+03	50	2.43370	0.472261	6.20118	0.917777	5.153297	6.756738
SEXRATIO (Scheduled Caste)	967	936.0	1089	949	977	153	1.089210E+03	33	2.42453	0.472261	7.55366	0.917777	5.133888	8.230385

SEXRATIO (Scheduled Tribe)	957	927.0	1020	942	965	93	5.877971E+02	24	1.20253	0.472261	1.09596	0.917777	2.546335	1.194151
Household Size	5	4.5	5	5	5	1	8.796106E-02	0	-0.06958	0.472261	-1.08187	0.917777	-0.14733	-1.1788
Household Size per square kilometre	0	0.0	0	0	0	0	6.659832E-08	0	0.45737	0.472261	-0.91418	0.917777	0.968467	-0.99608
Percentage of Scheduled Caste	11	1.4	19	9	14	18	2.111732E+01	5	-0.39831	0.472261	-0.09235	0.917777	-0.84341	-0.10063
Percentage of Scheduled Tribe	11	0.6	66	2	13	65	2.175097E+02	15	2.59965	0.472261	7.82422	0.917777	5.504695	8.525187
Infant Sex Ratio (0 - 6 yrs)	920	839.0	966	899	943	127	9.813460E+02	31	-0.59420	0.472261	0.26426	0.917777	-1.25821	0.287939
Male Literacy Rate	84	66.2	90	82	89	24	3.381685E+01	6	-1.77819	0.472261	3.58576	0.917777	-3.76526	3.907004
Female Literacy Rate	63	45.2	77	60	70	32	7.798652E+01	9	-0.46118	0.472261	-0.44794	0.917777	-0.97655	-0.48807
GENDER_GAP	21	12.8	30	19	24	17	1.994969E+01	4	-0.06470	0.472261	0.07459	0.917777	-0.137	0.08127
Work Participation Rate	44	37.8	51	42	47	13	9.411307E+00	3	-0.20587	0.472261	0.14172	0.917777	-0.43593	0.15442
Education Facilities	1134	651.0	1814	811	1443	1163	1.340420E+05	366	0.28192	0.472261	-1.15339	0.917777	0.596962	-1.25672
Education Facilities per square kilometre	0	0.1	0	0	0	0	6.213345E-04	0	-0.16569	0.472261	0.48750	0.917777	-0.35085	0.53118
Primary School	1128	650.0	1809	809	1432	1159	1.321288E+05	363	0.29784	0.472261	-1.10981	0.917777	0.630662	-1.20924
Primary School per square kilometre	0	0.1	0	0	0	0	6.249237E-04	0	-0.23199	0.472261	0.43406	0.917777	-0.49123	0.472944
Middle School	578	38.0	1068	368	729	1030	7.102188E+04	266	0.08148	0.472261	-0.50525	0.917777	0.172538	-0.55051
Middle School per	0	0.0	0	0	0	0	6.532781E-04	0	0.11971	0.472261	1.46021	0.917777	0.253476	1.591031

square kilometre														
Secondary School	253	91.0	559	158	295	468	1.623304E+04	127	1.07451	0.472261	0.35604	0.917777	2.275249	0.387938
Secondary School per square kilometre	0	0.0	0	0	0	0	1.041410E-04	0	0.72012	0.472261	1.49040	0.917777	1.524836	1.623928
Senior Secondary School	55	12.0	126	39	72	114	6.717971E+02	26	0.85542	0.472261	1.03479	0.917777	1.811332	1.1275
Senior Secondary School per square kilometre	0	0.0	0	0	0	0	9.217467E-06	0	0.56988	0.472261	-0.26249	0.917777	1.206704	-0.286
College	10	4.0	19	6	14	15	2.555797E+01	5	0.20024	0.472261	-1.18982	0.917777	0.423996	-1.29641
College per square kilometre	0	0.0	0	0	0	0	4.906146E-07	0	1.04059	0.472261	0.49196	0.917777	2.203426	0.536037
Adult Literacy centre	287	3.0	1329	49	434	1326	1.080678E+05	329	1.73412	0.472261	3.30868	0.917777	3.671956	3.605099
Adult Literacy centre per square kilometre	0	0.0	0	0	0	0	1.523328E-03	0	1.77639	0.472261	2.37609	0.917777	3.76146	2.588958
Training School	9	2.0	64	4	9	62	1.471141E+02	12	4.47097	0.472261	21.12088	0.917777	9.467171	23.01309
Training School per square kilometre	0	0.0	0	0	0	0	6.239116E-07	0	3.71252	0.472261	16.02443	0.917777	7.861167	17.46005
Industrial School	10	3.0	22	6	12	19	2.295652E+01	5	0.80854	0.472261	0.44383	0.917777	1.712068	0.483591
Industrial School per square kilometre	0	0.0	0	0	0	0	1.992334E-07	0	-0.10622	0.472261	-1.04428	0.917777	-0.22492	-1.13783
Literate Population_PER	74	55.8	84	71	79	28	5.018636E+01	7	-0.97319	0.472261	0.72613	0.917777	-2.06071	0.79118

Percentage of Male Literates	84	66.2	90	82	89	24	3.381685E+01	6	-1.77819	0.472261	3.58576	0.917777	-3.76526	3.907004
Percentage of Female Literates	63	45.2	77	60	70	32	7.798652E+01	9	-0.46118	0.472261	-0.44794	0.917777	-0.97655	-0.48807
Percentage of Total Workers	44	37.8	51	42	47	13	9.411307E+00	3	-0.20587	0.472261	0.14172	0.917777	-0.43593	0.15442
Percentage of Male Workers	52	48.9	57	51	54	8	3.852543E+00	2	0.12512	0.472261	-0.20281	0.917777	0.264944	-0.22098
Percentage of Female Workers	36	22.8	47	32	39	24	3.316468E+01	6	-0.56018	0.472261	0.04512	0.917777	-1.18617	0.049166
Percentage of Total Main Workers	36	26.9	42	34	39	15	1.103200E+01	3	-0.48226	0.472261	0.85725	0.917777	-1.02117	0.93405
Percentage of Male Main Workers	46	38.6	52	45	48	14	9.019713E+00	3	-0.59815	0.472261	0.69448	0.917777	-1.26657	0.756701
Percentage of Female Main Workers	25	16.0	35	22	28	19	2.201393E+01	5	0.19917	0.472261	0.16030	0.917777	0.421734	0.174657
Percentage of Total Marginal Workers	9	4.0	20	6	11	16	1.843058E+01	4	1.22479	0.472261	0.63980	0.917777	2.59346	0.697119
Percentage of Male Marginal Workers	6	3.1	16	4	7	13	1.149640E+01	3	1.53049	0.472261	1.87268	0.917777	3.240767	2.040447
Percentage of Female Marginal Workers	11	4.9	23	7	14	18	2.759662E+01	5	0.95210	0.472261	-0.08175	0.917777	2.016045	-0.08908
Percentage of Total Non Workers	56	48.8	62	53	58	13	9.411307E+00	3	0.20587	0.472261	0.14172	0.917777	0.435926	0.15442

Percentage of Male Non Workers	48	43.4	51	46	49	8	3.852543E+00	2	-0.12512	0.472261	-0.20281	0.917777	-0.26494	-0.22098
Percentage of Female Non Workers	64	53.1	77	61	68	24	3.316468E+01	6	0.56018	0.472261	0.04512	0.917777	1.186168	0.049166
Bicycle (num of households)	156297	29940.0	657541	58780	178375	627601	2.191906E+10	148051	2.21776	0.472261	5.34906	0.917777	4.696047	5.828279
Bicycle (Proportion)	32	10.9	59	20	44	49	2.223484E+02	15	0.59290	0.472261	-0.79753	0.917777	1.255445	-0.86898
Car, Jeep, Van (num of households)	11779	1165.0	94241	2728	12853	93076	3.626186E+08	19043	3.84261	0.472261	16.50091	0.917777	8.136636	17.97922
Car, Jeep, Van (proportion)	2	0.6	7	1	2	6	1.700580E+00	1	2.15746	0.472261	5.68100	0.917777	4.568373	6.189952
Television (Num of households)	171908	32964.0	911858	61968	221553	878894	3.539712E+10	188141	2.95584	0.472261	10.46452	0.917777	6.25892	11.40203
Television (proportion)	33	16.0	63	24	40	47	1.419069E+02	12	0.98572	0.472261	0.78886	0.917777	2.087237	0.859535
Permanent Houses (num)	216937	41959.0	1136863	76517	262082	1094904	5.351831E+10	231340	2.96198	0.472261	10.89503	0.917777	6.271918	11.8711
Permanent Houses (%)	42	16.8	79	33	53	62	2.483956E+02	16	0.73210	0.472261	0.12690	0.917777	1.550213	0.138264
Number of occupied census houses	517935	60330.0	1762107	284481	656887	1701777	1.248940E+11	353404	2.01108	0.472261	5.75013	0.917777	4.258419	6.265275
Proportion of occupied census houses	92	85.2	95	91	94	10	6.072174E+00	2	-0.95833	0.472261	0.96217	0.917777	-2.02923	1.048375

Number of hospitals, dispensaries etc	1813	474.0	7686	752	2466	7212	2.706729E+06	1645	2.30296	0.472261	6.38529	0.917777	4.876463	6.957346
Proportion of hospitals, dispensaries etc	0	0.2	0	0	0	0	4.275362E-03	0	0.68339	0.472261	-0.42424	0.917777	1.447055	-0.46225
Number of Schools, colleges etc	4483	275.0	12360	2975	5613	12085	5.744422E+06	2397	1.40570	0.472261	4.08246	0.917777	2.976538	4.4482
Proportion of Schools, Colleges etc	129	0.6	3065	1	1	3064	3.912118E+05	625	4.89898	0.472261	23.99999	0.917777	10.37346	26.15013
Electricity (num)	334528	89780.0	1264643	172111	419557	1174863	6.615257E+10	257201	2.27600	0.472261	6.72763	0.917777	4.819367	7.330348
Electricity (%)	71	43.6	93	60	81	49	1.904376E+02	14	-0.13516	0.472261	-0.80457	0.917777	-0.28619	-0.87665
Kerosene (Num)	110672	12720.0	274373	77163	137787	261653	3.271405E+09	57196	1.10030	0.472261	2.13802	0.917777	2.329865	2.329569
Kerosene (%)	28	6.7	55	18	38	48	1.806861E+02	13	0.13003	0.472261	-0.86089	0.917777	0.275337	-0.93802
Any Other (num)	2550	598.0	5047	1497	3404	4449	1.494653E+06	1223	0.24014	0.472261	-0.82436	0.917777	0.508486	-0.89821
Any Other (%)	1	0.3	1	0	1	1	8.717391E-02	0	0.88317	0.472261	-0.62977	0.917777	1.870084	-0.68619
No source (num)	2347	277.0	6487	1297	2539	6210	2.805028E+06	1675	1.30139	0.472261	1.06904	0.917777	2.755666	1.16481
No source (%)	1	0.1	3	0	1	3	2.182609E-01	0	3.90729	0.472261	17.59809	0.917777	8.273595	19.17469
urban Population per square kilometre	72	4.7	269	34	86	264	4.614136E+03	68	2.11841	0.472261	4.34077	0.917777	4.485686	4.729656
PCDDP (Rs 1998-99)	16651	11789.0	28878	13825	17633	17089	1.915030E+07	4376	1.67453	0.472261	2.94372	0.917777	3.545777	3.207442

Table 10. 2: Nature of distribution

Normal/Multi-peak distributions	**Asymmetrical/Skewed Distribution**
Number of Households	Population (Scheduled Caste)
Number of Households per square kilometre	Male (Scheduled Caste)
Total Population	female (Scheduled Caste)
Total Population per square kilometre	Population (Scheduled Tribe)
Total Male per square kilometre	Male (Scheduled Tribe)
Total Female per square kilometre	Female (Scheduled Tribe)
Literate Population per square kilometre	Total Male
Male Literates per square kilometre	Total Female
Female Literates per square kilometre	Total Population (0 - 6 yrs)
Illiterate Population	Male (0 - 6 yrs)
Male Illiterates	Female (0 - 6 yrs)
Female Illiterates	Literate Population
Total Workers per square kilometre	Male Literates
Total Male Workers per square kilometre	Female Literates
Total Female Workers per square	Total Workers

kilometre	
Total Main workers per square kilometre	Total Male Workers
Total Male Main workers per square kilometre	Total Female Workers
Total Female Main workers per square kilometre	Total Main workers
Marginal Workers (Persons)	Total Male Main workers
Marginal Workers (Persons)per square kilometre	Total Female Main workers
Marginal Workers (Male)	Household Size
Marginal Workers (Male)per square kilometre	Percentage of Scheduled Caste
Marginal Workers (Female)	Percentage of Scheduled Tribe
Marginal Workers (Female)per square kilometre	Male Literacy Rate
Non Workers (Persons)	Female Literacy Rate
Non Workers (Persons)per square kilometre	Training School
Non Workers (Male)	Training School per square kilometre
Non Workers (Male)per square kilometre	Percentage of Male Literates

Non Workers (Female)	Percentage of Female Literates
Non Workers (Female)per square kilometre	Bicycle (number of households)
SEXRATIO	CarJeepVan (number of households)
SEXRATIO (Scheduled Caste)	Television (number of households)
SEXRATIO (Scheduled Tribe)	Permanent Houses
Household Size per Area	Number of occupied census houses
Infant Sex Ratio (0 - 6 yrs)	Proportion of occupied census houses
GENDER_GAP	Number of hospitals, dispensaries etc
Work Participation Rate	Proportion of hospitals, dispensaries etc
Education Facilities	Proportion of Schools, Colleges etc
Education Facilities per square kilometre	
Primary School	
Primary School per square kilometre	
Middle School	
Middle School per square kilometre	

Secondary School	
Secondary School per square kilometre	
Senior Secondary School	
Senior Secondary School per square kilometre	
College	
College per Area	
Adult Literacy centre	
Adult Literacy centre per square kilometre	
Industrial School	
Industrial School per square kilometre	
Literate Population	
Total Population	

Percentage of Total Workers	
Percentage of Male Workers	
Percentage of Female Workers	
Percentage of Total Main Workers	
Percentage of Male Main Workers	
Percentage of Female Main Workers	
Percentage of Total Marginal Workers	
Percentage of Male Marginal Workers	
Percentage of Female Marginal Workers	
Percentage of Total Non Workers	
Percentage of Male Non Workers	
Percentage of Female Non Workers	

Bicycle (Proportion)	
Car, Jeep, Van (proportion)	
Television (proportion)	
Permanent Houses	
Number of Schools, colleges etc	
Electricity number	
Electricity	
Kerosene Number	
Kerosene	
Any Other number	
Any Other	
No source number	
No source	

Urban population per square Kilometre	
PCDDP (Rs1998 – 99)	

Table 10.3: Intercalibration information for all images

X-File	Y-File	Polynomial Power	Coefficient0	Coefficient1	Coefficient2	r-square	N_pts
F101992.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-1.9949	1.5821	-0.00898768	0.896655	36034
F101993.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.953216	1.58343	-0.009051	0.931575	39207
F101994.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.249302	1.47874	-0.00779456	0.921603	36763
F121994.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.647905	1.17382	-0.002462	0.906139	34580
F121995.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.037166	1.22305	-0.00368514	0.91549	37834
F121996.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.0593544	1.27183	-0.00402306	0.930866	36023
F121997.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.314912	1.17686	-0.00260719	0.924158	37698
F121998.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.0361421	1.06496	-0.00128982	0.952491	38094
F121999.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	3.86E-15	1	-9.05E-18	1	39471
F141997.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-1.11777	1.76592	-0.0121435	0.90958	37110
F141998.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.166583	1.6319	-0.0100748	0.971488	36999
F141999.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.146546	1.50343	-0.00775093	0.971528	39207
F142000.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	1.11989	1.31543	-0.00527415	0.927083	38157
F142001.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.259861	1.31752	-0.00505666	0.943024	38836
F142002.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	1.09042	1.18516	-0.00351602	0.918015	37201

F142003.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.817664	1.23578	-0.00387673	0.940848	38992
F152000.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.143311	1.04609	-0.00104032	0.930971	39140
F152001.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	-0.672933	1.10817	-0.00125711	0.958288	38915
F152002.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.0916143	0.955246	0.00101603	0.964125	38293
F152003.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.304656	1.50551	-0.00785383	0.929003	39090
F152004.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.617361	1.3276	-0.00501788	0.946358	37252
F152005.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.704214	1.27693	-0.00402004	0.931201	39203
F152006.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.831112	1.27934	-0.00406492	0.938434	38958
F152007.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	1.42026	1.29205	-0.00443158	0.900075	37231
F152008.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.658412	0.952268	5.91E-05	0.888378	26954
F162004.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.344563	1.18826	-0.00332935	0.901897	37094
F162005.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.0364735	1.41061	-0.00618441	0.937245	39294
F162006.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.107591	1.13815	-0.00163503	0.919398	37468
F162007.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.767322	0.90232	0.0015061	0.948348	38051
F162008.v4b.sicily.stable_lights.avg_vis	F121999.v4b.sicily.stable_lights.avg_vis	2	0.602385	0.989168	5.96E-05	0.94366	37757

Table 10.4: Available single orbit fixed gain radiance calibrated images for 2001

F15200101091732	F15200101091914	F15200101092056	F15200109130205
F15200109130347	F15200109130528	F15200109130710	F15200109130852
F15200109131034	F15200109131216	F15200109131357	F15200109131539
F15200109131721	F15200109131903	F15200109132045	F15200109132227
F15200109140008	F15200109140150	F15200109140332	F15200109140514
F15200109140656	F15200109140837	F15200109141019	F15200109141201
F15200109141343	F15200109141525	F15200109141707	F15200109141848

F15200109142030	F15200109142212	F15200109142354	F15200109150136
F15200109150317	F15200109150459	F15200109150641	F15200109150823
F15200109151005	F15200109151146	F15200109151328	F15200109151510
F15200109151652	F15200109151834	F15200109152016	F15200109152157
F15200109152339	F15200109160121	F15200109160303	F15200109160445
F15200109160626	F15200109160808	F15200109160950	F15200109161132
F15200109161314	F15200109161456	F15200109161637	F15200109161819
F15200109162001	F15200109162143	F15200109162325	F15200109170106
F15200109170248	F15200109170430	F15200109170612	F15200109170754
F15200109170936	F15200109171117	F15200109171259	F15200109171441
F15200109171623	F15200109171805	F15200109171946	F15200109172128
F15200109172310	F15200109180052	F15200109180234	F15200110120331
F15200110120513	F15200110120654	F15200110120836	F15200110121018
F15200110121200	F15200110121342	F15200110121523	F15200110121705
F15200110121847	F15200110122029	F15200110122211	F15200110122353
F15200110130134	F15200110130316	F15200110130640	F15200110130822
F15200110131003	F15200110131145	F15200110131327	F15200110131509
F15200110131651	F15200110131832	F15200110132014	F15200110132156
F15200110132338	F15200110140120	F15200110140302	F15200110140443
F15200110140625	F15200110140807	F15200110140949	F15200110141131
F15200110141312	F15200110141454	F15200110141636	F15200110141818
F15200110142000	F15200110142141	F15200110142323	F15200110150105
F15200110150247	F15200111110259	F15200111110440	F15200111110622

F15200111110804	F15200111110946	F15200111111128	F15200111111309
F15200111111451	F15200111111633	F15200111111815	F15200111111957
F15200111112138	F15200111112320	F15200111120102	F15200111120244
F15200111120426	F15200111120607	F15200111120749	F15200111120931
F15200111121113	F15200111121255	F15200111121437	F15200111121618
F15200111121800	F15200111121942	F15200111122124	F15200111122306
F15200111130047	F15200111130229	F15200111130411	F15200111130553
F15200111130735	F15200111130916	F15200111131058	F15200111131240
F15200111131422	F15200111131604	F15200111131745	F15200111131927
F15200111132109	F15200111132251	F15200111140033	F15200111140214
F15200111140356	F15200111140538	F15200111140720	F15200111140902
F15200111141044	F15200111141225	F15200111141407	F15200111141549
F15200111141731	F15200111141913	F15200111142054	F15200111142236
F15200111150018	F15200111150200	F15200111150342	F15200111150523
F15200111150705	F15200111150847	F15200111151029	F15200111151211
F15200111151352	F15200111151534	F15200111151716	F15200111151858
F15200111152040	F15200111152221	F15200111160003	F15200111160145
F15200111160327	F15200111170312	F15200111170454	F15200111170636
F15200111170818	F15200111170959	F15200111171141	F15200111171323
F15200111171505	F15200111171647	F15200111171828	F15200111172010
F15200111172152	F15200111172334	F15200111180116	F15200111180257
F15200111180439	F15200111180621	F15200111180803	F15200111180945
F15200111181127	F15200111181308	F15200111181450	F15200111181632

F15200111181814	F15200111181956	F15200111182137	F15200111182319
F15200111190101	F15200111190243	F15200111190425	F15200111190606
F15200111190748	F15200111190930	F15200111191112	F15200111191254
F15200111191435	F15200111191617	F15200111191759	F15200111191941
F15200111192123	F15200111192304	F15200111200046	F15200111200228

Table 10.5: Radiance per pixel information for F15 satellite as obtained from NGDC/NOAA

Gain dB	Saturation	Radiance
	radiance	per DN
	w/cm^2/sr	w/cm^2/sr
0	4.80E-06	7.62E-08
1	4.28E-06	6.79E-08
2	3.81E-06	6.05E-08
3	3.40E-06	5.39E-08
4	3.03E-06	4.81E-08
5	2.70E-06	4.28E-08
6	2.41E-06	3.82E-08
7	2.14E-06	3.40E-08
8	1.91E-06	3.03E-08
9	1.70E-06	2.70E-08
10	1.52E-06	2.41E-08
11	1.35E-06	2.15E-08
12	1.21E-06	1.91E-08
13	1.07E-06	1.71E-08
14	9.58E-07	1.52E-08

15	8.54E-07	1.35E-08
16	7.61E-07	1.21E-08
17	6.78E-07	1.08E-08
18	6.04E-07	9.59E-09
19	5.39E-07	8.55E-09
20	4.80E-07	7.62E-09
21	4.28E-07	6.79E-09
22	3.81E-07	6.05E-09
23	3.40E-07	5.39E-09
24	3.03E-07	4.81E-09
25	2.70E-07	4.28E-09
26	2.41E-07	3.82E-09
27	2.14E-07	3.40E-09
28	1.91E-07	3.03E-09
29	1.70E-07	2.70E-09
30	1.52E-07	2.41E-09
31	1.35E-07	2.15E-09
32	1.21E-07	1.91E-09
33	1.07E-07	1.71E-09
34	9.58E-08	1.52E-09
35	8.54E-08	1.35E-09
36	7.61E-08	1.21E-09
37	6.78E-08	1.08E-09
38	6.04E-08	9.59E-10
39	5.39E-08	8.55E-10

40	4.80E-08	7.62E-10
41	4.28E-08	6.79E-10
42	3.81E-08	6.05E-10
43	3.40E-08	5.39E-10
44	3.03E-08	4.81E-10
45	2.70E-08	4.28E-10
46	2.41E-08	3.82E-10
47	2.14E-08	3.40E-10
48	1.91E-08	3.03E-10
49	1.70E-08	2.70E-10
50	1.52E-08	2.41E-10
51	1.35E-08	2.15E-10
52	1.21E-08	1.91E-10
53	1.07E-08	1.71E-10
54	9.58E-09	1.52E-10
55	8.54E-09	1.35E-10
56	7.61E-09	1.21E-10
57	6.78E-09	1.08E-10
58	6.04E-09	9.59E-11
59	5.39E-09	8.55E-11
60	4.80E-09	7.62E-11
61	4.28E-09	6.79E-11
62	3.81E-09	6.05E-11
63	3.40E-09	5.39E-11
64	3.03E-09	4.81E-11

65	2.70E-09	4.28E-11

Table 10.6: Z-scores of Skewness and Kurtosis

Census Metrics	Skewness	Standard error	Kurtosis	Standard error	Z_Score_Skewness	Z_Score _Kurtosis
Other source of power supply (percentage)	0.88317	0.472261	-0.62977	0.917777	1.870084	-0.68619
Other source of power supply (number of households)	0.24014		-0.82436		0.508486	-0.89821
Percentage of households using bicycle	0.59290		-0.79753		1.255445	-0.86898
Number of Colleges	0.20024		-1.18982		0.423996	-1.29641
Number of Education Facilities	0.28192		-1.15339		0.596962	-1.25672
Number of Education Facilities per square kilometre	-0.16569		0.48750		-0.35085	0.53118
Percentage of households with Electricity as source of power	-0.13516		-0.80457		-0.28619	-0.87665
Percentage of female literates	-0.46118		-0.44794		-0.97655	-0.48807

Female literacy rate	-0.46118		-0.44794		-0.97655	-0.48807
Gender Gap	-0.06470		0.07459		-0.137	0.08127
Household size	-0.06958		-1.08187		-0.14733	-1.1788
Household size per square kilometre	0.45737		-0.91418		0.968467	-0.99608
Number of Industrial Schools	0.80854		0.44383		1.712068	0.483591
Number of Industrial Schools per square kilometre	-0.10622		-1.04428		-0.22492	-1.13783
Percentage of households with Kerosene as source of power	0.13003		-0.86089		0.275337	-0.93802
Main female agricultural labourers	0.63845		0.25020		1.351901	0.272612
Main agricultural labourers (persons)	0.77349		0.38462		1.637838	0.419083
Marginal Female Agricultural Labourers	0.36792		0.29708		0.779069	0.323696
Marginal Male Agricultural Labourers	0.72379		0.08189		1.532607	0.089222
Marginal Agricultural Labourers (persons)	0.49583		0.32318		1.049899	0.352132
Marginal Male Casual Labourers	0.79397		-0.76164		1.68122	-0.82988
Marginal Male Household Labourers	0.90614		0.28801		1.918725	0.313812
Female marginal workers	0.10838		-0.58384		0.229491	-0.63614
Female marginal workers per square kilometre	0.80211		-0.31277		1.698452	-0.34079
Male marginal workers	0.70952		0.56738		1.502396	0.618209
Marginal workers (persons)	0.14592		-0.60990		0.30899	-0.66454

Marginal workers (persons) per square kilometre	0.86573		0.00501		1.833159	0.005463
Number of Middle Schools	0.08148		-0.50525		0.172538	-0.55051
Number of Middle Schools per square kilometre	0.11971		1.46021		0.253476	1.591031
Male non-workers per square kilometre	0.62587		0.95676		1.32527	1.042472
Total non-workers per square kilometre	0.91238		1.23979		1.931935	1.350865
Percentage of female non-workers	0.56018		0.04512		1.186168	0.049166
Percentage of male non-workers	-0.12512		-0.20281		-0.26494	-0.22098
Percentage of total non-workers	0.20587		0.14172		0.435926	0.15442
Percentage of literate persons	-0.97319		0.72613		-2.06071	0.79118
Percentage of permanent houses	0.73210		0.12690		1.550213	0.138264
Number Primary School	0.29784		-1.10981		0.630662	-1.20924
Number of Primary School per square kilometre	-0.23199		0.43406		-0.49123	0.472944
Percentage of hospitals, and dispensaries	0.68339		-0.42424		1.447055	-0.46225
Percentage of occupied census houses	-0.95833		0.96217		-2.02923	1.048375
Percentage of Scheduled caste population	-0.39831		-0.09235		-0.84341	-0.10063
Number of Secondary Schools per	0.72012		1.49040		1.524836	1.623928

square kilometre						
Number of Senior Secondary Schools	0.85542		1.03479		1.811332	1.1275
Number of Senior Secondary Schools per square kilometre	0.56988		-0.26249		1.206704	-0.286
Infant Gender Ratio (0 – 6 yrs)	-0.59420		0.26426		-1.25821	0.287939
Total number of females per square kilometre	0.86121		1.56278		1.823587	1.702784
Total number of males per square kilometre	0.91309		1.41112		1.933452	1.537546
Percentage of total female main workers	0.19917		0.16030		0.421734	0.174657
Percentage of total male main workers	-0.59815		0.69448		-1.26657	0.756701
Percentage of total main workers	-0.48226		0.85725		-1.02117	0.93405
Total population per square kilometre	0.89601		1.49679		1.89728	1.630881
Percentage of total female workers	-0.56018		0.04512		-1.18617	0.049166
Percentage of total male workers	0.12512		-0.20281		0.264944	-0.22098
Percentage of total workers	-0.20587		0.14172		-0.43593	0.15442
Work Participation Rate	-0.20587		0.14172		-0.43593	0.15442

Table 10. 7 Shapiro – Wilks statistic

Census Metrics	**Shapiro-Wilk**	
	Statistic	**Level of Significance**
Other source of power supply (number of households)	.973	.741

Percentage of households using bicycle	.918	.054
Number of Education Facilities	.932	.107
Number of Education Facilities per square kilometre	.962	.475
Percentage of households with Electricity as source of power	.969	.647
Percentage of female literates	.961	.460
Female Literacy Rate	.961	.460
Gender Gap	.963	.492
Household size	.959	.415
Household size per square kilometre	.933	.117
Number of Industrial School	.935	.127
Number of Industrial School per square kilometre	.951	.278
Percentage of households with Kerosene as power source	.969	.647
Number of households with Kerosene as power source	.928	.088
Main female agricultural labourers	.951	.282
Main male agricultural labourers	.928	.090
Main agricultural labourers (persons)	.940	.161
Main female workers per square kilometre	.920	.059
Marginal female agricultural labourers	.968	.621
Marginal male agricultural labourers	.948	.243
Marginal agricultural labourers(persons)	.965	.537
Marginal workers (female)	.946	.218
Marginal workers (male)	.957	.373
Marginal workers (persons)	.961	.460

Number of Middle Schools	.976	.808
Number of Middle Schools per square kilometre	.927	.086
Male non-workers per square kilometre	.927	.085
Percentage of female non-workers	.961	.462
Percentage of male non-workers	.980	.905
Percentage of non-workers (persons)	.971	.681
Percentage of literate persons	.933	.117
Number of Permanent Houses	.934	.118
Number of Primary School	.934	.118
Number of Primary School square kilometre	.964	.517
Percentage of occupied census houses	.927	.082
Percentage of Scheduled Caste population	.971	.681
Number of Secondary Schools per square kilometre	.956	.365
Number of Senior Secondary Schools	.954	.331
Number of Senior Secondary School per square kilometre	.953	.310
Infant Gender Ratio (0 – 6 yrs)	.961	.452
Percentage of households with television	.923	.067
Percentage of main workers (female)	.977	.824
Percentage of main workers (male)	.963	.495
Percentage of main workers (persons)	.970	.662
Percentage of total workers (female)	.961	.462
Percentage of total workers (male)	.980	.905
Percentage of total workers (persons)	.971	.681

Work Participation Rate	.971	.681

Table 10.8: Bootstrap table of 144 variables

	mean brightness	Standard deviation brightness	Avg_DN	SD_DN
Number of households	0.83	0.66	0.75	0.50
Number of households per square kilometre	0.79	0.64	0.80	0.58
Total Population	0.83	0.66	0.76	0.53
Total population per square kilometre	0.80	0.64	0.83	0.63
Total number of males	0.83	0.67	0.77	0.53
Total Males per square Kilometre	0.81	0.65	0.84	0.64
Total number of females	0.83	0.66	0.76	0.52
Total Females per square Kilometre	0.78	0.62	0.82	0.61
Population (0 – 6 years)	0.81	0.65	0.76	0.54
Males (0 – 6 years)	0.81	0.64	0.76	0.54
Females (0 – 6 years)	0.82	0.66	0.76	0.54
Population (Scheduled Caste)	0.82	0.74	0.75	0.57
Males (Scheduled Caste)	0.82	0.74	0.76	0.58
Females (Scheduled Caste)	0.82	0.74	0.75	0.57
Population (Scheduled Tribe)	0.15	0.19	0.01	0.06
Males (Scheduled Tribe)	0.16	0.19	0.02	0.07
Females (Scheduled Tribe)	0.14	0.18	0.00	0.06
Literate population	0.85	0.69	0.77	0.53

Number of literates per square Kilometre	0.81	0.67	0.80	0.60
Male Literates	0.85	0.68	0.77	0.53
Male Literates per square Kilometre	0.81	0.66	0.82	0.62
Female Literates	0.85	0.70	0.76	0.52
Female Literates per square Kilometre	0.80	0.67	0.77	0.56
Illiterate population	0.74	0.56	0.71	0.49
Male Illiterates	0.75	0.58	0.71	0.50
Female Illiterates	0.72	0.55	0.71	0.49
Total Working population	0.78	0.59	0.73	0.48
Total Workers per square Kilometre	0.68	0.49	0.76	0.54
Total Male Workers	0.82	0.65	0.76	0.52
Total Male Workers per square Kilometre	0.79	0.61	0.82	0.61
Total Female Workers	0.65	0.44	0.66	0.40
Total female Workers per square Kilometre	0.37	0.18	0.54	0.33
Main Workers (Persons)	0.79	0.59	0.76	0.51
Main Workers (Persons) per square Kilometre	0.75	0.53	0.84	0.63
Main Workers (Male)	0.82	0.64	0.77	0.53
Main Workers (Male) per square Kilometre	0.81	0.61	0.86	0.65
Main Workers (Female)	0.67	0.46	0.70	0.45
Main Workers (Female) per square Kilometre	0.53	0.29	0.72	0.50
Main Casual Labourers (Persons)	0.51	0.28	0.58	0.33
Main Casual Labourers (Male)	0.52	0.28	0.60	0.37
Main Casual Labourers (Female)	0.48	0.27	0.53	0.27

Main Agricultural Labourers (Persons)	0.28	0.15	0.40	0.37
Main Agricultural Labourers (Male)	0.24	0.13	0.36	0.36
Main Agricultural Labourers (Female)	0.31	0.17	0.43	0.37
Main Household Industry Workers (Persons)	0.49	0.36	0.50	0.30
Main Household Industry Workers (Male)	0.69	0.51	0.67	0.42
Main Household Industry Workers (Female)	0.26	0.20	0.30	0.16
Main Other Workers (Persons)	0.87	0.73	0.74	0.49
Main Other Workers (Male)	0.88	0.74	0.75	0.50
Main Other Workers (Female)	0.84	0.69	0.70	0.44
Marginal workers (persons)	0.50	0.41	0.39	0.16
Marginal workers (persons) per square Kilometre	-0.04	0.00	-0.06	-0.15
Marginal workers (Male)	0.64	0.61	0.45	0.23
Marginal workers (Male) per square Kilometre	0.10	0.20	0.01	-0.09
Marginal workers (Female)	0.38	0.26	0.32	0.11
Marginal workers (Female) per square Kilometre	-0.12	-0.12	-0.09	-0.17
Marginal Casual Labourers (Persons)	0.14	0.00	0.10	-0.12
Marginal Casual Labourers (Male)	0.15	0.05	0.04	-0.19
Marginal Casual Labourers (Female)	0.13	-0.01	0.12	-0.09
Marginal Agricultural Labourers (Persons)	0.17	0.19	0.10	0.07
Marginal Agricultural Labourers (Male)	0.11	0.19	0.01	0.01
Marginal Agricultural Labourers (Female)	0.19	0.19	0.15	0.10
Marginal Household Industry Workers (Persons)	0.40	0.30	0.40	0.21
Marginal Household Industry Workers (Male)	0.57	0.47	0.48	0.24

Marginal Household Industry Workers (Female)	0.34	0.25	0.36	0.20
Marginal Other Workers (Persons)	0.86	0.77	0.70	0.45
Marginal Other Workers (Male)	0.88	0.83	0.70	0.47
Marginal Other Workers (Female)	0.76	0.61	0.65	0.39
Non Workers (Persons)	0.86	0.71	0.78	0.55
Non Workers (Persons) per square Kilometre	0.86	0.72	0.86	0.67
Non Workers (Male)	0.84	0.68	0.77	0.55
Non Workers (Male) per square Kilometre	0.82	0.68	0.86	0.67
Non Workers (Female)	0.87	0.72	0.78	0.55
Non Workers (Female) per square Kilometre	0.87	0.74	0.85	0.67
Gender ratio	-0.43	-0.34	-0.51	-0.51
Gender ratio (Scheduled Caste)	-0.43	-0.33	-0.57	-0.62
Gender ratio (Scheduled Tribe)	-0.27	-0.14	-0.39	-0.35
Household Size	0.07	0.01	0.26	0.37
Household Size per square Kilometre	-0.41	-0.34	-0.29	-0.19
Percentage of Scheduled Caste	0.20	0.26	0.25	0.25
Percentage of Scheduled Tribe	-0.20	-0.13	-0.34	-0.24
Gender ratio (0 – 6 years)	-0.33	-0.10	-0.55	-0.42
Male Literacy Rate	0.30	0.28	0.36	0.28
Female Literacy Rate	0.42	0.40	0.38	0.31
Gender Gap	-0.44	-0.43	-0.28	-0.25
Work Participation Rate	-0.72	-0.71	-0.71	-0.75
Education Facilities	0.46	0.36	0.37	0.20

Education Facilitiesper square kilometre	-0.13	-0.12	-0.05	-0.08
Primary School	0.47	0.36	0.39	0.22
Primary Schoolper square kilometre	-0.12	-0.11	-0.03	-0.05
Middle School	0.40	0.25	0.41	0.21
Middle Schoolper square kilometre	-0.02	-0.08	0.06	-0.04
Secondary School	0.45	0.23	0.50	0.28
Secondary Schoolper square kilometre	0.10	-0.09	0.26	0.10
Senior Secondary School	0.35	0.18	0.40	0.28
Senior Secondary Schoolper square kilometre	-0.06	-0.13	0.07	0.05
College	0.46	0.25	0.46	0.36
Collegeper square kilometre	0.08	-0.04	0.19	0.21
Adult Literacy centre	-0.19	-0.17	-0.28	-0.31
Adult Literacy centreper square kilometre	-0.33	-0.24	-0.34	-0.36
Training School	0.10	0.06	0.12	0.03
Training Schoolper square kilometre	0.03	-0.01	0.07	0.00
Industrial School	0.65	0.51	0.52	0.34
Industrial Schoolper square kilometre	0.29	0.22	0.28	0.22
Percentage of Literates (Persons)	0.40	0.38	0.40	0.32
Percentage of Male Literates	0.30	0.28	0.36	0.28
Percentage of Female Literates	0.42	0.40	0.38	0.31
Percentage of Total Workers (Persons)	-0.72	-0.71	-0.71	-0.75
Percentage of Total Workers (Male)	-0.04	-0.12	-0.15	-0.29
Percentage of Total Workers (Female)	-0.81	-0.76	-0.76	-0.75

Percentage of Total Main Workers (Persons)	0.02	-0.17	0.28	0.21
Percentage of Total Main Workers (Male)	0.47	0.21	0.65	0.55
Percentage of Total Main Workers (Female)	-0.34	-0.44	-0.10	-0.13
Percentage of Total Marginal Workers (Persons)	-0.53	-0.37	-0.72	-0.70
Percentage of Total Marginal Workers (Male)	-0.44	-0.25	-0.67	-0.66
Percentage of Total Marginal Workers (Female)	-0.58	-0.44	-0.74	-0.71
Percentage of non – workers (Persons)	0.72	0.71	0.71	0.75
Percentage of non – workers (Male)	0.04	0.12	0.15	0.29
Percentage of non – workers (Female)	0.81	0.76	0.76	0.75
Bicycle (num of households)	0.83	0.73	0.68	0.43
Bicycle (Proportion)	0.28	0.36	0.12	0.00
Car, Jeep, Van (num of households)	0.79	0.62	0.65	0.39
Car, Jeep, Van (proportion)	0.81	0.65	0.69	0.42
Television (Num of households)	0.87	0.73	0.74	0.50
Television (proportion)	0.88	0.79	0.79	0.62
Permanent Houses (num)	0.80	0.62	0.71	0.44
Permanent Houses (%)	0.65	0.47	0.64	0.42
Number of occupied census houses	0.77	0.60	0.69	0.41
Proportion of occupied census houses	-0.50	-0.39	-0.43	-0.21
Number of hospitals, dispensaries etc	0.87	0.75	0.76	0.53
Proportion of hospitals, dispensaries etc	0.78	0.73	0.61	0.41
Number of Schools, colleges etc	0.71	0.54	0.60	0.38
Proportion of Schools, Colleges etc	-0.21	-0.12	-0.38	-0.41

Electricity (num)	0.84	0.67	0.75	0.50
Electricity (%)	0.48	0.40	0.46	0.34
Kerosene (Num)	0.20	0.13	0.25	0.18
Kerosene (%)	-0.49	-0.41	-0.46	-0.35
Any Other (num)	0.54	0.44	0.57	0.49
Any Other (%)	-0.26	-0.22	-0.20	-0.08
No source (num)	0.66	0.60	0.54	0.45
No source (%)	-0.03	0.00	-0.07	0.03
urban Populationper square kilometre	0.96	0.89	0.82	0.65
PCDDP (Rs 1998-99)	0.81	0.81	0.55	0.34

Table 10.9: Amenities datasets for Maharashtra as obtained from the Indian Census, 2001

Indicator	**Data Value**	**Unit**	**Subpop**
Banking services - number of households	4842219	Number	Urban
Banking services - number of households	9175825	Number	Total
Banking services - number of households	4333606	Number	Rural
Banking services - proportion	48.1	Per cent	Total
Banking services - proportion	60	Per cent	Urban
Banking services - proportion	39.4	Per cent	Rural
Bathroom facility within the house - number of households	11651554	Number	Total
Bathroom facility within the house - number of households	5066823	Number	Rural
Bathroom facility within the house - number of households	6584731	Number	Urban

Bathroom facility within the house - proportion	46.1	Per cent	Rural
Bathroom facility within the house - proportion	61.1	Per cent	Total
Bathroom facility within the house - proportion	81.6	Per cent	Urban
Bicycle - number of households	3112816	Number	Rural
Bicycle - number of households	2618985	Number	Urban
Bicycle - number of households	5731801	Number	Total
Bicycle - proportion	32.5	Per cent	Urban
Bicycle - proportion	30.1	Per cent	Total
Bicycle - proportion	28.3	Per cent	Rural
Car, jeep, van - number of households	464567	Number	Urban
Car, jeep, van - number of households	642064	Number	Total
Car, jeep, van - number of households	177497	Number	Rural
Car, jeep, van - proportion	5.8	Per cent	Urban
Car, jeep, van - proportion	1.6	Per cent	Rural
Car, jeep, van - proportion	3.4	Per cent	Total
Closed drainage within the house - number of households	3637125	Number	Urban
Closed drainage within the house - number of households	4202901	Number	Total
Closed drainage within the house - number of households	565776	Number	Rural
Closed drainage within the house - proportion	45.1	Per cent	Urban
Closed drainage within the house - proportion	22	Per cent	Total
Closed drainage within the house - proportion	5.1	Per cent	Rural
Cooking in open - number of households	81276	Number	Urban
Cooking in open - number of households	531041	Number	Total

Cooking in open - number of households	449765	Number	Rural
Cooking in open - proportion	1	Per cent	Urban
Cooking in open - proportion	4.1	Per cent	Rural
Cooking in open - proportion	2.8	Per cent	Total
Drinking water source outside premises - Any other - number of households	97585	Number	Urban
Drinking water source outside premises - Any other - number of households	332543	Number	Total
Drinking water source outside premises - Any other - number of households	234958	Number	Rural
Drinking water source outside premises - Any other - proportion	4.5	Per cent	Urban
Drinking water source outside premises - Any other - proportion	3.7	Per cent	Total
Drinking water source outside premises - Any other - proportion	3.5	Per cent	Rural
Drinking water source outside premises - Handpump & Tube Well - number of households	2036217	Number	Rural
Drinking water source outside premises - Handpump & Tube Well - number of households	2345218	Number	Total
Drinking water source outside premises - Handpump & Tube Well - number of households	309001	Number	Urban
Drinking water source outside premises - Handpump & Tube Well - proportion	30.3	Per cent	Rural
Drinking water source outside premises - Handpump & Tube Well - proportion	14.3	Per cent	Urban
Drinking water source outside premises - Handpump & Tube Well - proportion	26.4	Per cent	Total
Drinking water source outside premises - number of households	2158812	Number	Urban
Drinking water source outside premises - number of households	8880756	Number	Total
Drinking water source outside premises - number of households	6721944	Number	Rural
Drinking water source outside premises - Proportion	46.6	Per cent	Total

Drinking water source outside premises - Proportion	26.8	Per cent	Urban
Drinking water source outside premises - Proportion	61.1	Per cent	Rural
Drinking water source outside premises - Tap - number of households	1615735	Number	Urban
Drinking water source outside premises - Tap - number of households	3575461	Number	Total
Drinking water source outside premises - Tap - number of households	1959726	Number	Rural
Drinking water source outside premises - Tap - proportion	74.8	Per cent	Urban
Drinking water source outside premises - Tap - proportion	40.3	Per cent	Total
Drinking water source outside premises - Tap - proportion	29.2	Per cent	Rural
Drinking water source outside premises - Well - number of households	132182	Number	Urban
Drinking water source outside premises - Well - number of households	2400394	Number	Rural
Drinking water source outside premises - Well - number of households	2532576	Number	Total
Drinking water source outside premises - Well - proportion	35.7	Per cent	Rural
Drinking water source outside premises - Well - proportion	6.1	Per cent	Urban
Drinking water source outside premises - Well - proportion	28.5	Per cent	Total
Drinking water source outside premises_Spring-number of households	4309	Number	Urban
Drinking water source outside premises_Spring-number of households	94958	Number	Total
Drinking water source outside premises_Spring-number of households	90649	Number	Rural
Drinking water source outside premises_Spring-proportion	0.2	Per cent	Urban
Drinking water source outside premises_Spring-proportion	1.1	Per cent	Total
Drinking water source outside premises_Spring-proportion	1.3	Per cent	Rural
Drinking water source within premises - proportion	73.2	Per cent	Urban
Drinking water source within premises - proportion	38.9	Per cent	Rural
Drinking water source within premises - proportion	53.4	Per cent	Total

Drinking water source within premises-Any other - number of households	12144	Number	Urban
Drinking water source within premises-Any other - number of households	29601	Number	Total
Drinking water source within premises-Any other - number of households	17457	Number	Rural
Drinking water source within premises-Any other - proportion	0.3	Per cent	Total
Drinking water source within premises-Any other - proportion	0.4	Per cent	Rural
Drinking water source within premises-Any other - proportion	0.2	Per cent	Urban
Drinking water source within premises-Handpump & tube well - number of households	478460	Number	Rural
Drinking water source within premises-Handpump & tube well - number of households	667569	Number	Total
Drinking water source within premises-Handpump & tube well - number of households	189109	Number	Urban
Drinking water source within premises-Handpump & tube well - proportion	3.2	Per cent	Urban
Drinking water source within premises-Handpump & tube well - proportion	6.6	Per cent	Total
Drinking water source within premises-Handpump & tube well - proportion	11.2	Per cent	Rural
Drinking water source within premises-Number of households	5910714	Number	Urban
Drinking water source within premises-Number of households	10182393	Number	Total
Drinking water source within premises-Number of households	4271679	Number	Rural
Drinking water source within premises-Tap - number of households	3047003	Number	Rural
Drinking water source within premises-Tap - number of households	8628031	Number	Total
Drinking water source within premises-Tap - number of households	5581028	Number	Urban
Drinking water source within premises-Tap - proportion	94.4	Per cent	Urban
Drinking water source within premises-Tap - proportion	84.7	Per cent	Total

Drinking water source within premises-Tap - proportion	71.3	Per cent	Rural
Drinking water source within premises-Well number of households	128433	Number	Urban
Drinking water source within premises-Well number of households	857192	Number	Total
Drinking water source within premises-Well number of households	728759	Number	Rural
Drinking water source within premises-Well - proportion	17.1	Per cent	Rural
Drinking water source within premises-Well - proportion	2.2	Per cent	Urban
Drinking water source within premises-Well - proportion	8.4	Per cent	Total
Drinking water source-Any other - number of households	109729	Number	Urban
Drinking water source-Any other - number of households	362144	Number	Total
Drinking water source-Any other - number of households	252415	Number	Rural
Drinking water source-Any other - proportion	1.4	Per cent	Urban
Drinking water source-Any other - proportion	1.9	Per cent	Total
Drinking water source-Any other - proportion	2.3	Per cent	Rural
Drinking water source-Handpump & tube Well - number of households	3012787	Number	Total
Drinking water source-Handpump & tube Well - number of households	498110	Number	Urban
Drinking water source-Handpump & tube Well - number of households	2514677	Number	Rural
Drinking water source-Handpump & tube Well - proportion	6.2	Per cent	Urban
Drinking water source-Handpump & tube Well - proportion	15.8	Per cent	Total
Drinking water source-Handpump & tube Well - proportion	22.9	Per cent	Rural
Drinking water source-Tap - number of households	7196763	Number	Urban
Drinking water source-Tap - number of households	12203492	Number	Total
Drinking water source-Tap - number of households	5006729	Number	Rural
Drinking water source-Tap - proportion	45.5	Per cent	Rural

Drinking water source-Tap - proportion	64	Per cent	Total
Drinking water source-Tap - proportion	89.2	Per cent	Urban
Drinking water source-Total-number of households	10993623	Number	Rural
Drinking water source-Total-number of households	8069526	Number	Urban
Drinking water source-Total-number of households	19063149	Number	Total
Drinking water source-Total-Spring-number of households	4309	Number	Urban
Drinking water source-Total-Spring-number of households	94958	Number	Total
Drinking water source-Total-Spring-number of households	90649	Number	Rural
Drinking water source-Total-Spring-proportion	0.5	Per cent	Total
Drinking water source-Total-Spring-proportion	0.1	Per cent	Urban
Drinking water source-Total-Spring-proportion	0.8	Per cent	Rural
Drinking water source-Well - number of households	260615	Number	Urban
Drinking water source-Well - number of households	3389768	Number	Total
Drinking water source-Well - number of households	3129153	Number	Rural
Drinking water source-Well - proportion	3.2	Per cent	Urban
Drinking water source-Well - proportion	17.8	Per cent	Total
Drinking water source-Well - proportion	28.5	Per cent	Rural
Five rooms - Number of Households	237904	Number	Total
Five rooms - Number of Households	137284	Number	Urban
Five rooms - Number of Households	100620	Number	Rural
Five rooms - proportion	1.7	Per cent	Urban
Five rooms - proportion	1.2	Per cent	Total
Five rooms - proportion	0.9	Per cent	Rural

Floor material-Any other - number of households	181012	Number	Urban
Floor material-Any other - number of households	273150	Number	Total
Floor material-Any other - number of households	92138	Number	Rural
Floor material-Any other - Proportion	2.2	Per cent	Urban
Floor material-Any other - Proportion	1.4	Per cent	Total
Floor material-Any other - Proportion	0.8	Per cent	Rural
Floor material-Brick - number of households	78151	Number	Rural
Floor material-Brick - number of households	132516	Number	Total
Floor material-Brick - number of households	54365	Number	Urban
Floor material-Brick - Proportion	0.7	Per cent	Rural
Floor material-Brick - Proportion	0.7	Per cent	Urban
Floor material-Brick - Proportion	0.7	Per cent	Total
Floor material-Cement - number of households	1845315	Number	Urban
Floor material-Cement - number of households	2602306	Number	Total
Floor material-Cement - number of households	756991	Number	Rural
Floor material-Cement - Proportion	13.7	Per cent	Total
Floor material-Cement - Proportion	22.9	Per cent	Urban
Floor material-Cement - Proportion	6.9	Per cent	Rural
Floor material-Mosaic, floor tiles - number of households	4450698	Number	Urban
Floor material-Mosaic, floor tiles - number of households	5586344	Number	Total
Floor material-Mosaic, floor tiles - number of households	1135646	Number	Rural
Floor material-Mosaic, floor tiles - Proportion	55.2	Per cent	Urban
Floor material-Mosaic, floor tiles - Proportion	29.3	Per cent	Total

Floor material-Mosaic, floor tiles - Proportion	10.3	Per cent	Rural
Floor material-Mud - number of households	1122397	Number	Urban
Floor material-Mud - number of households	8530531	Number	Rural
Floor material-Mud - number of households	9652928	Number	Total
Floor material-Mud - Proportion	77.6	Per cent	Rural
Floor material-Mud - Proportion	13.9	Per cent	Urban
Floor material-Mud - Proportion	50.6	Per cent	Total
Floor material-Stone - number of households	401464	Number	Urban
Floor material-Stone - number of households	774502	Number	Total
Floor material-Stone - number of households	373038	Number	Rural
Floor material-Stone - Proportion	5	Per cent	Urban
Floor material-Stone - Proportion	4.1	Per cent	Total
Floor material-Stone - Proportion	3.4	Per cent	Rural
Floor material-Wood, bamboo - number of households	14275	Number	Urban
Floor material-Wood, bamboo - number of households	27128	Number	Rural
Floor material-Wood, bamboo - number of households	41403	Number	Total
Floor material-Wood, bamboo - Proportion	0.2	Per cent	Urban
Floor material-Wood, bamboo - Proportion	0.2	Per cent	Total
Floor material-Wood, bamboo - Proportion	0.2	Per cent	Rural
Four rooms - Number of Households	434534	Number	Urban
Four rooms - Number of Households	823976	Number	Total
Four rooms - Number of Households	389442	Number	Rural
Four rooms - proportion	5.4	Per cent	Urban

Four rooms - proportion	4.3	Per cent	Total
Four rooms - proportion	3.5	Per cent	Rural
Fuel for cooking - Any other (including electricity & bio-gas) - number of households	22914	Number	Urban
Fuel for cooking - Any other (including electricity & bio-gas) - number of households	162124	Number	Total
Fuel for cooking - Any other (including electricity & bio-gas) - number of households	139210	Number	Rural
Fuel for cooking - Any other (including electricity & bio-gas) - proportion	0.9	Per cent	Total
Fuel for cooking - Any other (including electricity & bio-gas) - proportion	1.3	Per cent	Rural
Fuel for cooking - Any other (including electricity & bio-gas) - proportion	0.3	Per cent	Urban
Fuel for cooking - Coal, lignite, charcoal - number of households	13494	Number	Rural
Fuel for cooking - Coal, lignite, charcoal - number of households	53044	Number	Total
Fuel for cooking - Coal, lignite, charcoal - number of households	39550	Number	Urban
Fuel for cooking - Coal, lignite, charcoal - proportion	0.5	Per cent	Urban
Fuel for cooking - Coal, lignite, charcoal - proportion	0.3	Per cent	Total
Fuel for cooking - Coal, lignite, charcoal - proportion	0.1	Per cent	Rural
Fuel for cooking - Cowdung cake - number of households	27785	Number	Urban
Fuel for cooking - Cowdung cake - number of households	400225	Number	Total
Fuel for cooking - Cowdung cake - number of households	372440	Number	Rural
Fuel for cooking - Cowdung cake - proportion	3.4	Per cent	Rural
Fuel for cooking - Cowdung cake - proportion	2.1	Per cent	Total
Fuel for cooking - Cowdung cake - proportion	0.3	Per cent	Urban

Fuel for cooking - Crop residue - number of households	96804	Number	Urban
Fuel for cooking - Crop residue - number of households	939241	Number	Total
Fuel for cooking - Crop residue - number of households	842437	Number	Rural
Fuel for cooking - Crop residue - proportion	1.2	Per cent	Urban
Fuel for cooking - Crop residue - proportion	4.9	Per cent	Total
Fuel for cooking - Crop residue - proportion	7.7	Per cent	Rural
Fuel for cooking - Firewood - number of households	8076135	Number	Rural
Fuel for cooking - Firewood - number of households	802354	Number	Urban
Fuel for cooking - Firewood - number of households	8878489	Number	Total
Fuel for cooking - Firewood - proportion	9.9	Per cent	Urban
Fuel for cooking - Firewood - proportion	46.6	Per cent	Total
Fuel for cooking - Firewood - proportion	73.5	Per cent	Rural
Fuel for cooking - Kerosene - number of households	2424717	Number	Urban
Fuel for cooking - Kerosene - number of households	2897008	Number	Total
Fuel for cooking - Kerosene - number of households	472291	Number	Rural
Fuel for cooking - Kerosene - proportion	15.2	Per cent	Total
Fuel for cooking - Kerosene - proportion	30	Per cent	Urban
Fuel for cooking - Kerosene - proportion	4.3	Per cent	Rural
Fuel for cooking - L.P.G. - number of households	4601342	Number	Urban
Fuel for cooking - L.P.G. - number of households	5656425	Number	Total
Fuel for cooking - L.P.G. - number of households	1055083	Number	Rural
Fuel for cooking - L.P.G. - proportion	57	Per cent	Urban
Fuel for cooking - L.P.G. - proportion	29.7	Per cent	Total

Fuel for cooking - L.P.G. - proportion	9.6	Per cent	Rural
Married couples with independent sleeping rooms-number	4524341	Number	Urban
Married couples with independent sleeping rooms-number	7087693	Number	Rural
Married couples with independent sleeping rooms-number	11612034	Number	Total
Married couples with independent sleeping rooms-proportion	58.4	Per cent	Rural
Married couples with independent sleeping rooms-proportion	54.7	Per cent	Urban
Married couples with independent sleeping rooms-proportion	56.9	Per cent	Total
No cooking at all - number of households	22533	Number	Rural
No cooking at all - number of households	76593	Number	Total
No cooking at all - number of households	54060	Number	Urban
No cooking at all - proportion	0.2	Per cent	Rural
No cooking at all - proportion	0.7	Per cent	Urban
No cooking at all - proportion	0.4	Per cent	Total
No drainage within the house - number of households	6470832	Number	Rural
No drainage within the house - number of households	7473121	Number	Total
No drainage within the house - number of households	1002289	Number	Urban
No drainage within the house - proportion	12.4	Per cent	Urban
No drainage within the house - proportion	39.2	Per cent	Total
No drainage within the house - proportion	58.9	Per cent	Rural
No exclusive room - Number of Households	538253	Number	Rural
No exclusive room - Number of Households	311518	Number	Urban
No exclusive room - Number of Households	849771	Number	Total
No exclusive room - proportion	3.9	Per cent	Urban

No exclusive room - proportion	4.5	Per cent	Total
No exclusive room - proportion	4.9	Per cent	Rural
No latrine within the house - number of households	3382994	Number	Urban
No latrine within the house - number of households	12374681	Number	Total
No latrine within the house - number of households	8991687	Number	Rural
No latrine within the house - proportion	81.8	Per cent	Rural
No latrine within the house - proportion	41.9	Per cent	Urban
No latrine within the house - proportion	64.9	Per cent	Total
No separate kitchen within the house - number of households	1288771	Number	Urban
No separate kitchen within the house - number of households	3322643	Number	Total
No separate kitchen within the house - number of households	2033872	Number	Rural
No separate kitchen within the house - proportion	17.4	Per cent	Total
No separate kitchen within the house - proportion	16	Per cent	Urban
No separate kitchen within the house - proportion	18.5	Per cent	Rural
No source of lighting - number of households	28162	Number	Urban
No source of lighting - number of households	90780	Number	Total
No source of lighting - number of households	62618	Number	Rural
No source of lighting - proportion	0.3	Per cent	Urban
No source of lighting - proportion	0.5	Per cent	Total
No source of lighting - proportion	0.6	Per cent	Rural
None of the specified assets - number of households	1497722	Number	Urban
None of the specified assets - number of households	7016634	Number	Total
None of the specified assets - number of households	5518912	Number	Rural

None of the specified assets - proportion	36.8	Per cent	Total
None of the specified assets - proportion	50.2	Per cent	Rural
None of the specified assets - proportion	18.6	Per cent	Urban
Number of occupied census houses	9799666	Number	Urban
Number of occupied census houses	22992781	Number	Total
Number of occupied census houses	13193115	Number	Rural
Number of factories	219248	Number	Urban
Number of factories	296706	Number	Total
Number of factories	77458	Number	Rural
Number of hospitals, dispensaries etc.	48879	Number	Urban
Number of hospitals, dispensaries etc.	30490	Number	Rural
Number of hospitals, dispensaries etc.	79369	Number	Total
Number of hotels, lodges, guest houses etc.	49620	Number	Urban
Number of hotels, lodges, guest houses etc.	87821	Number	Total
Number of hotels, lodges, guest houses etc.	38201	Number	Rural
Number of houses with other non-residential uses	1854373	Number	Total
Number of houses with other non-residential uses	1450375	Number	Rural
Number of houses with other non-residential uses	403998	Number	Urban
Number of partly residential houses	311765	Number	Rural
Number of partly residential houses	514619	Number	Total
Number of partly residential houses	202854	Number	Urban
Number of places of worship	51406	Number	Urban
Number of places of worship	220458	Number	Total

Number of places of worship	169052	Number	Rural
Number of schools, colleges etc.	32988	Number	Urban
Number of schools, colleges etc.	178118	Number	Total
Number of schools, colleges etc.	145130	Number	Rural
Number of shops, offices etc.	434613	Number	Rural
Number of shops, offices etc.	1455971	Number	Total
Number of shops, offices etc.	1021358	Number	Urban
Number of wholly residential census houses	7769315	Number	Urban
Number of wholly residential census houses	18305346	Number	Total
Number of wholly residential census houses	10536031	Number	Rural
Occupied by any other means - number of households	353964	Number	Urban
Occupied by any other means - number of households	378270	Number	Rural
Occupied by any other means - number of households	732234	Number	Total
Occupied by any other means - proportion	3.4	Per cent	Rural
Occupied by any other means - proportion	4.4	Per cent	Urban
Occupied by any other means - proportion	3.8	Per cent	Total
One room - Number of Households	3769120	Number	Urban
One room - Number of Households	9124234	Number	Total
One room - Number of Households	5355114	Number	Rural
One room - proportion	46.7	Per cent	Urban
One room - proportion	48.7	Per cent	Rural
One room - proportion	47.9	Per cent	Total
Open drainage within the house - number of households	3430112	Number	Urban

Open drainage within the house - number of households	7387127	Number	Total
Open drainage within the house - number of households	3957015	Number	Rural
Open drainage within the house - proportion	42.5	Per cent	Urban
Open drainage within the house - proportion	38.8	Per cent	Total
Open drainage within the house - proportion	36	Per cent	Rural
Other latrine within the house - number of households	827338	Number	Total
Other latrine within the house - number of households	535330	Number	Urban
Other latrine within the house - number of households	292008	Number	Rural
Other latrine within the house - proportion	6.6	Per cent	Urban
Other latrine within the house - proportion	4.3	Per cent	Total
Other latrine within the house - proportion	2.7	Per cent	Rural
Owned census houses - number of households	5419455	Number	Urban
Owned census houses - number of households	15310948	Number	Total
Owned census houses - number of households	9891493	Number	Rural
Owned census houses - proportion	67.2	Per cent	Urban
Owned census houses - proportion	80.3	Per cent	Total
Owned census houses - proportion	90	Per cent	Rural
Permanent houses-Number of households	4434160	Number	Rural
Permanent houses-Number of households	11021426	Number	Total
Permanent houses-Number of households	6587266	Number	Urban
Permanent houses-Proportion	40.3	Per cent	Rural
Permanent houses-Proportion	81.6	Per cent	Urban
Permanent houses-Proportion	57.8	Per cent	Total

Pit latrine within the house - number of households	571036	Number	Urban
Pit latrine within the house - number of households	1695494	Number	Total
Pit latrine within the house - number of households	1124458	Number	Rural
Pit latrine within the house - proportion	10.2	Per cent	Rural
Pit latrine within the house - proportion	8.9	Per cent	Total
Pit latrine within the house - proportion	7.1	Per cent	Urban
Proportion of partly residential houses	1.8	Per cent	Urban
Proportion of partly residential houses	2	Per cent	Total
Proportion of partly residential houses	2.2	Per cent	Rural
Proportion of factories	0.5	Per cent	Rural
Proportion of factories	2	Per cent	Urban
Proportion of factories	1.2	Per cent	Total
Proportion of hospitals, dispensaries etc.	0.4	Per cent	Urban
Proportion of hospitals, dispensaries etc.	0.3	Per cent	Total
Proportion of hospitals, dispensaries etc.	0.2	Per cent	Rural
Proportion of hotels, lodges, guest houses etc.	0.4	Per cent	Urban
Proportion of hotels, lodges, guest houses etc.	0.3	Per cent	Total
Proportion of hotels, lodges, guest houses etc.	0.3	Per cent	Rural
Proportion of houses with other non-residential uses	7.2	Per cent	Total
Proportion of houses with other non-residential uses	3.6	Per cent	Urban
Proportion of houses with other non-residential uses	10.1	Per cent	Rural
Proportion of occupied census houses	87.6	Per cent	Urban
Proportion of occupied census houses	89.8	Per cent	Total

Proportion of occupied census houses	91.6	Per cent	Rural
Proportion of place of worship to total census houses	0.5	Per cent	Urban
Proportion of place of worship to total census houses	0.9	Per cent	Total
Proportion of place of worship to total census houses	1.2	Per cent	Rural
Proportion of schools, colleges etc.	1	Per cent	Rural
Proportion of schools, colleges etc.	0.7	Per cent	Total
Proportion of schools, colleges etc.	0.3	Per cent	Urban
Proportion of shops, offices etc.	3	Per cent	Rural
Proportion of shops, offices etc.	9.1	Per cent	Urban
Proportion of shops, offices etc.	5.7	Per cent	Total
Proportion of vacant census houses	12.4	Per cent	Urban
Proportion of vacant census houses	10.2	Per cent	Total
Proportion of vacant census houses	8.4	Per cent	Rural
Proportion of wholly residential census houses	71.5	Per cent	Total
Proportion of wholly residential census houses	69.4	Per cent	Urban
Proportion of wholly residential census houses	73.1	Per cent	Rural
Radio, transistor - number of households	2946401	Number	Rural
Radio, transistor - number of households	6843363	Number	Total
Radio, transistor - number of households	3896962	Number	Urban
Radio, transistor - proportion	48.3	Per cent	Urban
Radio, transistor - proportion	35.9	Per cent	Total
Radio, transistor - proportion	26.8	Per cent	Rural
Rented census houses - number of households	2296107	Number	Urban

Rented census houses - number of households	723860	Number	Rural
Rented census houses - number of households	3019967	Number	Total
Rented census houses - proportion	28.5	Per cent	Urban
Rented census houses - proportion	15.8	Per cent	Total
Rented census houses - proportion	6.6	Per cent	Rural
Roof material-Any other material - number of households	227528	Number	Total
Roof material-Any other material - number of households	152625	Number	Rural
Roof material-Any other material - number of households	74903	Number	Urban
Roof material-Any other material - proportion	1.4	Per cent	Rural
Roof material-Any other material - proportion	1.2	Per cent	Total
Roof material-Any other material - proportion	0.9	Per cent	Urban
Roof material-Brick - number of households	50426	Number	Urban
Roof material-Brick - number of households	93763	Number	Total
Roof material-Brick - number of households	43337	Number	Rural
Roof material-Brick - proportion	0.6	Per cent	Urban
Roof material-Brick - proportion	0.5	Per cent	Total
Roof material-Brick - proportion	0.4	Per cent	Rural
Roof material-Concrete - number of households	708105	Number	Rural
Roof material-Concrete - number of households	4030108	Number	Total
Roof material-Concrete - number of households	3322003	Number	Urban
Roof material-Concrete - proportion	41.2	Per cent	Urban
Roof material-Concrete - proportion	21.1	Per cent	Total
Roof material-Concrete - proportion	6.4	Per cent	Rural

Roof material-G.I., metal, asbestos sheets - number of households	3012662	Number	Urban
Roof material-G.I., metal, asbestos sheets - number of households	6638816	Number	Total
Roof material-G.I., metal, asbestos sheets - number of households	3626154	Number	Rural
Roof material-G.I., metal, asbestos sheets - proportion	33	Per cent	Rural
Roof material-G.I., metal, asbestos sheets - proportion	37.3	Per cent	Urban
Roof material-G.I., metal, asbestos sheets - proportion	34.8	Per cent	Total
Roof material-Grass, thatch, bamboo, wood, mud etc. - number of households	177843	Number	Urban
Roof material-Grass, thatch, bamboo, wood, mud etc. - number of households	1915626	Number	Total
Roof material-Grass, thatch, bamboo, wood, mud etc. - number of households	1737783	Number	Rural
Roof material-Grass, thatch, bamboo, wood, mud etc. - proportion	2.2	Per cent	Urban
Roof material-Grass, thatch, bamboo, wood, mud etc. - proportion	10	Per cent	Total
Roof material-Grass, thatch, bamboo, wood, mud etc. - proportion	15.8	Per cent	Rural
Roof material-Plastic, polythene - number of households	143032	Number	Total
Roof material-Plastic, polythene - number of households	94000	Number	Urban
Roof material-Plastic, polythene - number of households	49032	Number	Rural
Roof material-Plastic, polythene - proportion	1.2	Per cent	Urban
Roof material-Plastic, polythene - proportion	0.8	Per cent	Total
Roof material-Plastic, polythene - proportion	0.4	Per cent	Rural
Roof material-Slate - number of households	33028	Number	Urban
Roof material-Slate - number of households	67130	Number	Total
Roof material-Slate - number of households	34102	Number	Rural
Roof material-Slate - proportion	0.3	Per cent	Rural
Roof material-Slate - proportion	0.4	Per cent	Total

Roof material-Slate - proportion	0.4	Per cent	Urban
Roof material-Stone - number of households	22324	Number	Rural
Roof material-Stone - number of households	10469	Number	Urban
Roof material-Stone - number of households	32793	Number	Total
Roof material-Stone - proportion	0.1	Per cent	Urban
Roof material-Stone - proportion	0.2	Per cent	Total
Roof material-Stone - proportion	0.2	Per cent	Rural
Roof material-Tiles - number of households	5914353	Number	Total
Roof material-Tiles - number of households	1294192	Number	Urban
Roof material-Tiles - number of households	4620161	Number	Rural
Roof material-Tiles - proportion	16	Per cent	Urban
Roof material-Tiles - proportion	31	Per cent	Total
Roof material-Tiles - proportion	42	Per cent	Rural
Scooter, motor, cycle, moped - number of households	1635428	Number	Urban
Scooter, motor, cycle, moped - number of households	2513953	Number	Total
Scooter, motor, cycle, moped - number of households	878525	Number	Rural
Scooter, motor, cycle, moped - proportion	8	Per cent	Rural
Scooter, motor, cycle, moped - proportion	13.2	Per cent	Total
Scooter, motor, cycle, moped - proportion	20.3	Per cent	Urban
Semi-permanent houses-Number of households	1279213	Number	Urban
Semi-permanent houses-Number of households	6552768	Number	Total
Semi-permanent houses-Number of households	5273555	Number	Rural
Semi-permanent houses-Proportion	34.4	Per cent	Total

Semi-permanent houses-Proportion	15.9	Per cent	Urban
Semi-permanent houses-Proportion	48	Per cent	Rural
Separate kitchen within the house - number of households	6645419	Number	Urban
Separate kitchen within the house - number of households	15132872	Number	Total
Separate kitchen within the house - number of households	8487453	Number	Rural
Separate kitchen within the house - proportion	82.4	Per cent	Urban
Separate kitchen within the house - proportion	79.4	Per cent	Total
Separate kitchen within the house - proportion	77.2	Per cent	Rural
Six rooms and above - Number of Households	157956	Number	Urban
Six rooms and above - Number of Households	269018	Number	Total
Six rooms and above - Number of Households	111062	Number	Rural
Six rooms and above - proportion	1.4	Per cent	Total
Six rooms and above - proportion	1	Per cent	Rural
Six rooms and above - proportion	2	Per cent	Urban
Source of lighting - Any other - number of households	71567	Number	Rural
Source of lighting - Any other - number of households	24886	Number	Urban
Source of lighting - Any other - number of households	96453	Number	Total
Source of lighting - Any other - proportion	0.3	Per cent	Urban
Source of lighting - Any other - proportion	0.5	Per cent	Total
Source of lighting - Any other - proportion	0.7	Per cent	Rural
Source of lighting - Electricity - number of households	7608033	Number	Urban
Source of lighting - Electricity - number of households	14772090	Number	Total
Source of lighting - Electricity - number of households	7164057	Number	Rural

Source of lighting - Electricity - proportion	77.5	Per cent	Total
Source of lighting - Electricity - proportion	94.3	Per cent	Urban
Source of lighting - Electricity - proportion	65.2	Per cent	Rural
Source of lighting - Kerosene - number of households	408445	Number	Urban
Source of lighting - Kerosene - number of households	4103826	Number	Total
Source of lighting - Kerosene - number of households	3695381	Number	Rural
Source of lighting - Kerosene - proportion	5.1	Per cent	Urban
Source of lighting - Kerosene - proportion	21.5	Per cent	Total
Source of lighting - Kerosene - proportion	33.6	Per cent	Rural
Telephone - number of households	2204969	Number	Urban
Telephone - number of households	2686081	Number	Total
Telephone - number of households	481112	Number	Rural
Telephone - proportion	27.3	Per cent	Urban
Telephone - proportion	14.1	Per cent	Total
Telephone - proportion	4.4	Per cent	Rural
Television - number of households	2717806	Number	Rural
Television - number of households	5692391	Number	Urban
Television - number of households	8410197	Number	Total
Television - proportion	70.5	Per cent	Urban
Television - proportion	44.1	Per cent	Total
Television - proportion	24.7	Per cent	Rural
Temporary houses-Number of households	193919	Number	Urban
Temporary houses-Number of households	1475031	Number	Total

Temporary houses-Number of households	1281112	Number	Rural
Temporary houses-Proportion	2.4	Per cent	Urban
Temporary houses-Proportion	7.7	Per cent	Total
Temporary houses-Proportion	11.7	Per cent	Rural
Three rooms - Number of Households	1151251	Number	Rural
Three rooms - Number of Households	2240009	Number	Total
Three rooms - Number of Households	1088758	Number	Urban
Three rooms - proportion	13.5	Per cent	Urban
Three rooms - proportion	11.8	Per cent	Total
Three rooms - proportion	10.5	Per cent	Rural
Total number of vacant census houses	1392728	Number	Urban
Total number of vacant census houses	2608608	Number	Total
Total number of vacant census houses	1215880	Number	Rural
Total number of census houses	11192394	Number	Urban
Total number of census houses	25601389	Number	Total
Total number of census houses	14408995	Number	Rural
Total number of households	19063149	Number	Total
Total number of households	10993623	Number	Rural
Total number of households	8069526	Number	Urban
Total number of married couples	8265893	Number	Urban
Total number of married couples	20398163	Number	Total
Total number of married couples	12132270	Number	Rural
Two rooms - Number of Households	2170356	Number	Urban

Two rooms - Number of Households	5518237	Number	Total
Two rooms - Number of Households	3347881	Number	Rural
Two rooms - proportion	30.5	Per cent	Rural
Two rooms - proportion	28.9	Per cent	Total
Two rooms - proportion	26.9	Per cent	Urban
Unclassifiable houses-Number of households	9128	Number	Urban
Unclassifiable houses-Number of households	4796	Number	Rural
Unclassifiable houses-Number of households	13924	Number	Total
Unclassifiable houses-Proportion	0	Per cent	Rural
Unclassifiable houses-Proportion	0.1	Per cent	Urban
Unclassifiable houses-Proportion	0.1	Per cent	Total
Wall material-Any other - number of households	23350	Number	Urban
Wall material-Any other - number of households	55098	Number	Total
Wall material-Any other - number of households	31748	Number	Rural
Wall material-Any other - proportion	0.3	Per cent	Urban
Wall material-Any other - proportion	0.3	Per cent	Total
Wall material-Any other - proportion	0.3	Per cent	Rural
Wall material-Burnt brick - number of households	2786381	Number	Rural
Wall material-Burnt brick - number of households	5080753	Number	Urban
Wall material-Burnt brick - number of households	7867134	Number	Total
Wall material-Burnt brick - proportion	63	Per cent	Urban
Wall material-Burnt brick - proportion	41.3	Per cent	Total
Wall material-Burnt brick - proportion	25.3	Per cent	Rural

Wall material-Concrete - number of households	1132352	Number	Urban
Wall material-Concrete - number of households	1535711	Number	Total
Wall material-Concrete - number of households	403359	Number	Rural
Wall material-Concrete - proportion	14	Per cent	Urban
Wall material-Concrete - proportion	3.7	Per cent	Rural
Wall material-Concrete - proportion	8.1	Per cent	Total
Wall material-G.I., metal, asbestos sheets - number of households	321907	Number	Urban
Wall material-G.I., metal, asbestos sheets - number of households	390552	Number	Total
Wall material-G.I., metal, asbestos sheets - number of households	68645	Number	Rural
Wall material-G.I., metal, asbestos sheets - proportion	2	Per cent	Total
Wall material-G.I., metal, asbestos sheets - proportion	0.6	Per cent	Rural
Wall material-G.I., metal, asbestos sheets - proportion	4	Per cent	Urban
Wall material-Grass, thatch, bamboo etc. - number of households	1220130	Number	Rural
Wall material-Grass, thatch, bamboo etc. - number of households	1389608	Number	Total
Wall material-Grass, thatch, bamboo etc. - number of households	169478	Number	Urban
Wall material-Grass, thatch, bamboo etc. - proportion	2.1	Per cent	Urban
Wall material-Grass, thatch, bamboo etc. - proportion	7.3	Per cent	Total
Wall material-Grass, thatch, bamboo etc. - proportion	11.1	Per cent	Rural
Wall material-Mud, unburnt brick - number of households	1013403	Number	Urban
Wall material-Mud, unburnt brick - number of households	5629526	Number	Total
Wall material-Mud, unburnt brick - number of households	4616123	Number	Rural
Wall material-Mud, unburnt brick - proportion	42	Per cent	Rural
Wall material-Mud, unburnt brick - proportion	29.5	Per cent	Total

Wall material-Mud, unburnt brick - proportion	12.6	Per cent	Urban
Wall material-Plastic, polythene - number of households	63092	Number	Urban
Wall material-Plastic, polythene - number of households	93602	Number	Total
Wall material-Plastic, polythene - number of households	30510	Number	Rural
Wall material-Plastic, polythene - proportion	0.8	Per cent	Urban
Wall material-Plastic, polythene - proportion	0.5	Per cent	Total
Wall material-Plastic, polythene - proportion	0.3	Per cent	Rural
Wall material-Stone - number of households	1723856	Number	Rural
Wall material-Stone - number of households	136168	Number	Urban
Wall material-Stone - number of households	1860024	Number	Total
Wall material-Stone - proportion	1.7	Per cent	Urban
Wall material-Stone - proportion	9.8	Per cent	Total
Wall material-Stone - proportion	15.7	Per cent	Rural
Wall material-Wood - number of households	129023	Number	Urban
Wall material-Wood - number of households	241894	Number	Total
Wall material-Wood - number of households	112871	Number	Rural
Wall material-Wood - proportion	1.3	Per cent	Total
Wall material-Wood - proportion	1.6	Per cent	Urban
Wall material-Wood - proportion	1	Per cent	Rural
Water closet latrine within the house - number of households	585470	Number	Rural
Water closet latrine within the house - number of households	3580166	Number	Urban
Water closet latrine within the house - number of households	4165636	Number	Total
Water closet latrine within the house - proportion	44.4	Per cent	Urban

Water closet latrine within the house - proportion	21.9	Per cent	Total
Water closet latrine within the house - proportion	5.3	Per cent	Rural

Table 10.10: Demographic Datasets as obtained from the Indian census, 2001 (a)

NAME	Number of Households	Total Population	Total Male	Total Female	Total Population (0 - 6 yrs)	Male (0 - 6 yrs)	Female (0 - 6 yrs)	Population (Scheduled Caste)	Male (Scheduled Caste)	female (Scheduled Caste)	Population (Scheduled Tribe)	Male (Scheduled Tribe)	Female (Scheduled Tribe)
Nandurbar *	245421	1311709	663511	648198	230213	117386	112827	41412	21054	20358	859574	427858	431716
Dhule	324557	1707947	878372	829575	255231	133861	121370	109102	55756	53346	443564	224727	218837
Jalgaon	732767	3682690	1905493	1777197	525668	279551	246117	286777	146704	140073	435951	224017	211934
Buldana	445634	2232480	1147403	1085077	340294	178332	161962	241623	123418	118205	115156	59168	55988
Akola	319947	1630239	841253	788986	235835	122004	113831	168447	86365	82082	100088	51601	48487
Washim *	200958	1020216	526094	494122	160486	83688	76798	162663	83196	79467	70987	36439	34548
Amravati	526230	2607160	1345614	1261546	357834	184329	173505	446623	230641	215982	356533	182217	174316
Wardha	270502	1236736	638990	597746	155612	80704	74908	158630	81939	76691	154415	79668	74747
Nagpur	838599	4067637	2105314	1962323	525850	270823	255027	696461	357314	339147	444441	228119	216322
Bhandara	244531	1136146	573445	562701	154051	78749	75302	201949	101260	100689	97718	49059	48659
Gondiya *	249720	1200707	598834	601873	171191	87427	83764	167699	83317	84382	196455	97249	99206
Gadchiroli	210362	970294	491101	479193	154744	78724	76020	108824	54882	53942	371696	187017	184679
Chandrapur	462632	2071101	1062993	1008108	279490	144117	135373	296927	152141	144786	375256	190558	184698

Yavatmal	518214	2458271	1265681	1192590	369402	191135	178267	252802	129482	123320	473370	241500	231870
Nanded	527875	2876259	1481358	1394901	477303	247468	229835	498196	255571	242625	253596	129515	124081
Hingoli *	180695	987160	505373	481787	167098	86717	80381	100697	51036	49661	86898	44735	42163
Parbhani	278702	1527715	780191	747524	252435	131276	121159	152463	77373	75090	35210	17917	17293
Jalna	303886	1612980	826903	786077	261386	137345	124041	181017	92252	88765	32103	16380	15723
Aurangabad	549898	2897013	1505363	1391650	467934	247542	220392	376181	193907	182274	100416	51292	49124
Nashik	915137	4993796	2590912	2402884	789398	411061	378337	426516	218344	208172	1E+06	604271	590000
Thane	1755124	8131849	4377747	3754102	1E+06	592830	552066	339720	177990	161730	1E+06	600809	598481
Mumbai (Suburban) *	1838426	8640419	4741720	3898699	1E+06	532988	491712	401569	211010	190559	70454	37335	33119
Mumbai	677163	3338031	1878246	1459785	339723	176789	162934	183469	94811	88658	20666	11187	9479
Raigarh	478933	2207929	1117628	1090301	314767	162365	152402	53667	27112	26555	269124	135880	133244
Pune	1517041	7232555	3769128	3463427	968851	509367	459484	761857	390965	370892	261722	133835	127887
Ahmadnagar	776787	4040642	2083053	1957589	589706	312953	276753	484685	247545	237140	303255	153590	149665
Bid	414973	2161250	1116356	1044894	335283	177060	158223	281240	144022	137218	24193	12553	11640
Latur	381600	2080285	1075257	1005028	326777	170392	156385	404251	208233	196018	47836	24789	23047
Osmanabad	295750	1486586	769368	717218	223183	117861	105322	245790	126577	119213	27857	14468	13389
Solapur	735092	3849543	1989623	1859920	569609	300628	268981	578123	297283	280840	68989	35711	33278
Satara	570606	2808994	1408326	1400668	368531	196241	172290	246110	124214	121896	21896	11218	10678
Ratnagiri	377366	1696777	794498	902279	236601	121196	115405	24515	12254	12261	20102	10410	9692
Sindhudurg	192666	868825	417890	450935	105518	54277	51241	38536	18446	20090	4952	2515	2437
Kolhapur	712349	3523162	1807470	1715692	449883	244682	205201	449641	228078	221563	21387	10971	10416
Sangli	506593	2583524	1320088	1263436	341643	184564	157079	313474	158570	154904	17855	9176	8679

Table 10.11: Demographic Datasets as obtained from the Indian census, 2001 (b)

	P _LIT	Male Literates	Female Literates	Illiterate Population	Male Illiterates	Female Illiterates	TOT_ WORK _P	TOTAL _WORK _M	TOTAL _WORK _F	MAIN WORK _P	MAIN WORK _M	MAIN WORK _F	MAIN _CL _P	MAIN _CL_M
Nandurbar *	603221	361333	241888	708488	302178	406310	610033	344452	265581	458321	295678	162643	177962	119629
Dhule	1040837	606036	434801	667110	272336	394774	737941	454239	283702	573496	400309	173187	179514	124598
Jalgaon	2381376	1396860	984516	1301314	508633	792681	1563983	976955	587028	1300086	879454	420632	340002	224165
Buldana	1433889	842410	591479	798591	304993	493598	1016015	596607	419408	908797	557077	351720	325895	192469
Akola	1135245	639483	495762	494994	201770	293224	648243	435791	212452	582264	408232	174032	114058	84635
Washim *	630763	377978	252785	389453	148116	241337	455332	267742	187590	395651	247952	147699	131843	84831
Amravati	1856549	1032546	824003	750611	313068	437543	1095622	724019	371603	908363	648109	260254	185644	144235
Wardha	865556	486736	378820	371180	152254	218926	550351	348611	201740	431330	301631	129699	123306	87240
Nagpur	2976205	1654342	1321863	1091432	450972	640460	1538349	1091626	446723	1283079	969195	313884	199076	130849
Bhandara	770662	440122	330540	365484	133323	232161	536885	303036	233849	365464	237168	128296	105965	65897
Gondiya *	808414	458266	350148	392293	140568	251725	579596	319021	260575	391890	251939	139951	136153	85223
Gadchiroli	490121	296314	193807	480173	194787	285386	497104	272390	224714	336959	214170	122789	177806	110098
Chandrapur	1311008	762137	548871	760093	300856	459237	930791	567591	363200	683029	471479	211550	207577	132645
Yavatmal	1537777	903639	634138	920494	362042	558452	1118937	674923	444014	949453	612955	336498	284888	188136
Nanded	1625685	992485	633200	1250574	488873	761701	1202037	721922	480115	965087	642694	322393	339606	221830
Hingoli *	543309	337912	205397	443851	167461	276390	459978	259474	200504	412838	242942	169896	193689	109822
Parbhani	842536	516714	325822	685179	263477	421702	644752	384405	260347	559917	361626	198291	213367	132127
Jalna	870399	545730	324669	742581	281173	461408	716900	415137	301763	611998	383705	228293	292018	176025
Aurangabad	1771659	1067801	703858	1125354	437562	687792	1176857	743791	433066	1025531	686275	339256	400081	233316

Nashik	3126188	1823366	1302822	1867608	767546	1100062	2185573	1367857	817716	1911340	1265529	645811	747243	425089
Thane	5635799	3295251	2340548	2496050	1082496	1413554	3179981	2443327	736654	2782004	2257126	524878	293629	167350
Mumbai (Suburban) *	6617264	3853376	2763888	2023155	888344	1134811	3152509	2653135	499374	2971039	2524384	446655	2658	1803
Mumbai	2590613	1535176	1055437	747418	343070	404348	1311739	1111415	200324	1250710	1068656	182054	1289	847
Raigarh	1458324	822913	635411	749605	294715	454890	913995	588943	325052	664689	484273	180416	193279	105057
Pune	5039290	2879761	2159529	2193265	889367	1303898	2954482	2016053	938429	2645429	1883972	761457	709537	391397
Ahmadnagar	2598597	1517029	1081568	1442045	566024	876021	1857046	1095450	761596	1611061	1010920	600141	748820	425251
Bid	1241434	758028	483406	919816	358328	561488	957584	547915	409669	830658	504611	326047	424195	237612
Latur	1254534	750455	504079	825751	324802	500949	840518	525457	315061	712849	492299	220550	261904	176397
Osmanabad	872022	523930	348092	614564	245438	369126	653522	390202	263320	555202	358490	196712	236068	155780
Solapur	2336825	1384746	952079	1512718	604877	907841	1743041	1056182	686859	1485967	969288	516679	525775	331254
Satara	1908947	1069029	839918	900047	339297	560750	1303658	755834	547824	1044989	677712	367277	515217	301981
Ratnagiri	1095833	578330	517503	600944	216168	384776	763387	406780	356607	557613	337788	219825	292293	129548
Sindhudurg	612919	328199	284720	255906	89691	166215	404985	228357	176628	233827	161489	72338	106440	62843
Kolhapur	2364307	1367021	997286	1158855	440449	718406	1652970	1023027	629943	1394882	942612	452270	561840	332240
Sangli	1717836	979509	738327	865688	340579	525109	1218655	740429	478226	952241	664556	287685	433071	288765

Table 10.12: Demographic Datasets as obtained from the Indian census, 2001 (c)

	Female Main Casual	Total Main Agricultu	Male Main Agricultur	Female Main Agricultur	Total Main Household	Male Main Househol	Female Main Househol	Main Other Workers	Main Other Workers	Main Other Workers	Marginal Workers (Persons)	Marginal Workers (Male)	Marginal Workers (Female)	Marginal Casual Labourer

	Labourer	ral Labourer	al Labourer	al Labourer	Labourer	d Labourer	d Labourer	(Persons)	(Male)	(Female)				(Persons)
Nandurbar *	58333	180009	91662	88347	7331	4805	2526	93019	79582	13437	151712	48774	102938	27605
Dhule	54916	200046	106084	93962	14923	10391	4532	179013	159236	19777	164445	53930	110515	24812
Jalgaon	115837	546762	290286	256476	27258	18878	8380	386064	346125	39939	263897	97501	166396	36997
Buldana	133426	400535	202152	198383	10727	7071	3656	171640	155385	16255	107218	39530	67688	27766
Akola	29423	271692	150579	121113	6785	4531	2254	189729	168487	21242	65979	27559	38420	8773
Washim *	47012	194966	102055	92911	4573	3252	1321	64269	57814	6455	59681	19790	39891	11180
Amravati	41409	428625	248368	180257	14601	10362	4239	279493	245144	34349	187259	75910	111349	23718
Wardha	36066	157229	85131	72098	8257	6142	2115	142538	123118	19420	119021	46980	72041	15603
Nagpur	68227	241648	124839	116809	28424	19486	8938	813931	694021	119910	255270	122431	132839	26270
Bhandara	40068	125148	65367	59781	20562	9948	10614	113789	95956	17833	171421	65868	105553	28808
Gondiya *	50930	104727	58568	46159	38711	13495	25216	112299	94653	17646	187706	67082	120624	47045
Gadchiroli	67708	86273	43886	42387	5707	4043	1664	67173	56143	11030	160145	58220	101925	41493
Chandrapur	74932	190172	94979	95193	14827	11014	3813	270453	232841	37612	247762	96112	151650	31577
Yavatmal	96752	431234	224799	206435	10524	7614	2910	222807	192406	30401	169484	61968	107516	23447
Nanded	117776	349876	186102	163774	20461	11175	9286	255144	223587	31557	236950	79228	157722	46957
Hingoli *	83867	146822	71282	75540	5669	3909	1760	66658	57929	8729	47140	16532	30608	12745
Parbhani	81240	198438	99527	98911	7023	4873	2150	141089	125099	15990	84835	22779	62056	24247
Jalna	115993	179336	87277	92059	10129	6392	3737	130515	114011	16504	104902	31432	73470	25770
Aurangabad	166765	216776	104349	112427	14786	8906	5880	393888	339704	54184	151326	57516	93810	41044
Nashik	322154	406687	195579	211108	35127	21667	13460	722283	623194	99089	274233	102328	171905	76426

Thane	126279	176030	94793	81237	56543	36714	19829	2E+06	2E+06	297533	397977	186201	211776	95046
Mumbai (Suburban) *	855	2234	1631	603	78448	57373	21075	3E+06	2E+06	424122	181470	128751	52719	324
Mumbai	442	736	529	207	35316	25662	9654	1E+06	1E+06	171751	61029	42759	18270	143
Raigarh	88222	76067	41817	34250	15290	10732	4558	380053	326667	53386	249306	104670	144636	67701
Pune	318140	277732	133362	144370	61047	35752	25295	2E+06	1E+06	273652	309053	132081	176972	84047
Ahmadna gar	323569	358925	175864	183061	40705	22814	17891	462611	386991	75620	245985	84530	161455	75323
Bid	186583	213170	101340	111830	11851	7828	4023	181442	157831	23611	126926	43304	83622	38781
Latur	85507	242041	134694	107347	11472	7757	3715	197432	173451	23981	127669	33158	94511	38055
Osmanaba d	80288	196582	100154	96428	11207	6951	4256	111345	95605	15740	98320	31712	66608	20907
Solapur	194521	377394	186349	191045	69419	19815	49604	513379	431870	81509	257074	86894	170180	69664
Satara	213236	185229	88025	97204	24944	16563	8381	319599	271143	48456	258669	78122	180547	115079
Ratnagiri	162745	45128	24222	20906	9387	6976	2411	210805	177042	33763	205774	68992	136782	110268
Sindhudur g	43597	23963	15387	8576	7380	5180	2200	96044	78079	17965	171158	66868	104290	82552
Kolhapur	229600	217497	108234	109263	43690	28799	14891	571855	473339	98516	258088	80415	177673	104074
Sangli	144306	191490	102380	89110	26982	17016	9966	300698	256395	44303	266414	75873	190541	127320

Table 10.13: Demographic Datasets as obtained from the Indian census, 2001 (d)

	Marginal Casual Labourer	**Marginal Casual Labourer**	**Marginal Agricultural**	**Marginal Agricultural**	**Marginal Agricultural**	**Marginal Household**	**Marginal Household**	**Marginal Household**	**Marginal Other Worker**	**Marginal Other Worker**	**Marginal Other Worker**	**Non Workers (Persons)**	**Non Workers (Male)**	**Non Workers (Female)**

	(Male)	(Female)	Labourer (Persons)	Labourer (Male)	Labourer (Female)	Labourer (Persons)	Labourer (Male)	Labourer (Female)	(Persons)	(Male)	(Female)			
Nandurbar *	9394	18211	111913	33349	78564	2914	783	2131	9280	5248	4032	701676	319059	382617
Dhule	7866	16946	115822	33282	82540	6185	1536	4649	17626	11246	6380	970006	424133	545873
Jalgaon	11517	25480	184266	58887	125379	9552	2998	6554	33082	24099	8983	2E+06	928538	1E+06
Buldana	9290	18476	64628	20995	43633	2977	815	2162	11847	8430	3417	1E+06	550796	665669
Akola	3189	5584	41908	13995	27913	1955	538	1417	13343	9837	3506	981996	405462	576534
Washim *	3151	8029	43477	13529	29948	1168	331	837	3856	2779	1077	564884	258352	306532
Amravati	9944	13774	133538	46947	86591	4893	1446	3447	25110	17573	7537	2E+06	621595	889943
Wardha	5812	9791	81829	26878	54951	2977	1133	1844	18612	13157	5455	686385	290379	396006
Nagpur	10479	15791	127559	43876	83683	9659	3267	6392	91782	64809	26973	3E+06	1E+06	2E+06
Bhandara	11258	17550	118532	41716	76816	8224	2117	6107	15857	10777	5080	599261	270409	328852
Gondiya *	17559	29486	105738	35905	69833	18255	2866	15389	16668	10752	5916	621111	279813	341298
Gadchiroli	16920	24573	103143	32909	70234	2480	831	1649	13029	7560	5469	473190	218711	254479
Chandrapur	12642	18935	175452	57282	118170	5205	2219	2986	35528	23969	11559	1E+06	495402	644908
Yavatmal	7979	15468	124976	40289	84687	2811	1017	1794	18250	12683	5567	1E+06	590758	748576
Nanded	11772	35185	151504	47143	104361	8490	1745	6745	29999	18568	11431	2E+06	759436	914786
Hingoli *	3813	8932	28025	8819	19206	1082	341	741	5288	3559	1729	527182	245899	281283
Parbhani	4334	19913	46362	10184	36178	2010	470	1540	12216	7791	4425	882963	395786	487177
Jalna	5786	19984	59966	14925	45041	4121	806	3315	15045	9915	5130	896080	411766	484314
Aurangabad	12335	28709	72989	22023	50966	5524	1381	4143	31769	21777	9992	2E+06	761572	958584
Nashik	24598	51828	133415	41317	92098	12551	3143	9408	51841	33270	18571	3E+06	1E+06	2E+06

Thane	30894	64152	130613	49544	81069	23085	5917	17168	149233	99846	49387	5E+06	2E+06	3E+06
Mumbai (Suburban) *	181	143	481	265	216	16598	5293	11305	164067	123012	41055	5E+06	2E+06	3E+06
Mumbai	77	66	235	158	77	5980	2174	3806	54671	40350	14321	2E+06	766831	1E+06
Raigarh	23136	44565	110028	39887	70141	9008	3096	5912	62569	38551	24018	1E+06	528685	765249
Pune	27258	56789	96977	26325	70652	17702	4868	12834	110327	73630	36697	4E+06	2E+06	3E+06
Ahmadnagar	21808	53515	111257	31760	79497	13426	3252	10174	45979	27710	18269	2E+06	987603	1E+06
Bid	11996	26785	66913	18950	47963	3999	1088	2911	17233	11270	5963	1E+06	568441	635225
Latur	6136	31919	72472	17379	55093	3584	781	2803	13558	8862	4696	1E+06	549800	689967
Osmanabad	5669	15238	60543	17535	43008	4910	837	4073	11960	7671	4289	833064	379166	453898
Solapur	21894	47770	123583	32332	91251	14692	2671	12021	49135	29997	19138	2E+06	933441	1E+06
Satara	29827	85252	95784	23329	72455	11731	2954	8777	36075	22012	14063	2E+06	652492	852844
Ratnagiri	29613	80655	56817	17693	39124	5596	1924	3672	33093	19762	13331	933390	387718	545672
Sindhudurg	28262	54290	54930	20734	34196	6981	2235	4746	26695	15637	11058	463840	189533	274307
Kolhapur	29569	74505	75423	17891	57532	15413	3562	11851	63178	29393	33785	2E+06	784443	1E+06
Sangli	33754	93566	92945	24350	68595	13494	2540	10954	32655	15229	17426	1E+06	579659	785210

Table 10.14: Demographic Datasets as obtained from the Indian census, 2001 (e)

	SEXRATIO	SEXRATIO_Scheduled Caste	SEXRATIO_Scheduled Tribe	Household Size	Percentage of Scheduled	Percentage of Schedule	Infant Sex Ratio (0 - 6 yrs)	Male Literacy Rate	Female Literacy Rate	GENDER_GAP	Literate Population_PER	Percentage of Male Literates	Percentage of Female	Total Population_PER

					Caste	d Tribe							Literates	
Nandurbar *	977	967	1009	5.34473	3.1571	65.531	961	66.163	45.181	20.982	55.777	66.163	45.181	100
Dhule	944	957	974	5.26239	6.3879	25.971	907	81.401	61.395	20.006	71.648	81.401	61.395	100
Jalgaon	933	955	946	5.02573	7.78716	11.838	880	85.911	64.302	21.609	75.431	85.911	64.302	100
Buldana	946	958	946	5.00967	10.8231	5.1582	908	86.93	64.074	22.855	75.779	86.93	64.074	100
Akola	938	950	940	5.09534	10.3327	6.1395	933	88.91	73.429	15.48	81.414	88.91	73.429	100
Washim *	939	955	948	5.07676	15.944	6.958	918	85.437	60.573	24.864	73.368	85.437	60.573	100
Amravati	938	936	957	4.95441	17.1306	13.675	941	88.914	75.733	13.181	82.538	88.914	75.733	100
Wardha	935	936	938	4.572	12.8265	12.486	928	87.184	72.455	14.729	80.061	87.184	72.455	100
Nagpur	932	949	948	4.85051	17.122	10.926	942	90.18	77.424	12.756	84.031	90.18	77.424	100
Bhandara	981	994	992	4.64622	17.7749	8.6008	956	88.968	67.817	21.151	78.471	88.968	67.817	100
Gondiya *	1005	1013	1020	4.80821	13.9667	16.362	958	89.609	67.582	22.027	78.524	89.609	67.582	100
Gadchiroli	976	983	987	4.6125	11.2156	38.308	966	71.855	48.07	23.785	60.097	71.855	48.07	100
Chandrapur	948	952	969	4.47678	14.3367	18.119	939	82.942	62.891	20.051	73.175	82.942	62.891	100
Yavatmal	942	952	960	4.74374	10.2837	19.256	933	84.095	62.518	21.577	73.618	84.095	62.518	100
Nanded	942	949	958	5.44875	17.321	8.8169	929	80.435	54.349	26.087	67.766	80.435	54.349	100
Hingoli *	953	973	943	5.46313	10.2007	8.8028	927	80.714	51.169	29.544	66.252	80.714	51.169	100
Parbhani	958	970	965	5.48154	9.97981	2.3047	923	79.627	52.018	27.609	66.067	79.627	52.018	100
Jalna	951	962	960	5.30785	11.2225	1.9903	903	79.142	49.041	30.101	64.398	79.142	49.041	100
Aurangabad	924	940	958	5.26827	12.9851	3.4662	890	84.893	60.094	24.799	72.935	84.893	60.094	100
Nashik	927	953	976	5.45688	8.54092	23.915	920	83.646	64.351	19.295	74.355	83.646	64.351	100
Thane	858	909	996	4.6332	4.17765	14.748	931	87.063	73.096	13.967	80.662	87.063	73.096	100

Mumbai (Suburban) *	822	903	887	4.6999	4.64756	0.8154	923	91.557	81.124	10.433	86.89	91.557	81.124	100
Mumbai	777	935	847	4.92944	5.49632	0.6191	922	90.227	81.385	8.8425	86.402	90.227	81.385	100
Raigarh	976	979	981	4.6101	2.43065	12.189	939	86.145	67.748	18.397	77.031	86.145	67.748	100
Pune	919	949	956	4.76754	10.5337	3.6187	902	88.343	71.89	16.453	80.452	88.343	71.89	100
Ahmadnagar	940	958	974	5.20174	11.9952	7.5051	884	85.703	64.347	21.356	75.301	85.703	64.347	100
Bid	936	953	927	5.20817	13.0128	1.1194	894	80.702	54.519	26.183	67.988	80.702	54.519	100
Latur	935	941	930	5.45148	19.4325	2.2995	918	82.936	59.398	23.537	71.544	82.936	59.398	100
Osmanabad	932	942	925	5.0265	16.5339	1.8739	894	80.418	56.887	23.531	69.022	80.418	56.887	100
Solapur	935	945	932	5.23682	15.018	1.7921	895	81.986	59.844	22.143	71.246	81.986	59.844	100
Satara	995	981	952	4.92283	8.7615	0.7795	878	88.198	68.376	19.821	78.221	88.198	68.376	100
Ratnagiri	1136	1001	931	4.49637	1.4448	1.1847	952	85.895	65.767	20.128	75.048	85.895	65.767	100
Sindhudurg	1079	1089	969	4.50949	4.43542	0.57	944	90.261	71.234	19.026	80.298	90.261	71.234	100
Kolhapur	949	971	949	4.94584	12.7624	0.607	839	87.473	66.024	21.449	76.931	87.473	66.024	100
Sangli	957	977	946	5.0998	12.1336	0.6911	851	86.261	66.735	19.526	76.625	86.261	66.735	100

Table 10.15: Demographic Datasets as obtained from the Indian census, 2001 (f)

	Percentage of Total Workers	**Percentage of Male Workers**	**Percentage of Female Workers**	**Percentage of Total Main Workers**	**Percentage of Male Main Workers**	**Percentage of Female Main**	**Percentage of Total Marginal Workers**	**Percentage of Male Marginal Workers**	**Percentage of Female Marginal**	**Percentage of Total Non Workers**	**Percentage of Male Non Workers**	**Percentage of Female Non**	**CL_PER_P**	**CL_PER_M**

						Workers			Workers			Workers		
Nandurbar *	46.507	51.9135	40.9722	34.9408	44.5626	25.092	11.566	7.3509	15.881	53.493	48.086	59.028	33.698	37.457
Dhule	43.206	51.7137	34.1985	33.5781	45.574	20.877	9.6282	6.1398	13.322	56.794	48.286	65.802	27.689	29.162
Jalgaon	42.468	51.2705	33.0311	35.3026	46.1536	23.668	7.1659	5.1168	9.3628	57.532	48.73	66.969	24.105	24.124
Buldana	45.511	51.9963	38.6524	40.708	48.5511	32.414	4.8026	3.4452	6.2381	54.489	48.004	61.348	34.809	33.818
Akola	39.764	51.8026	26.9272	35.7165	48.5267	22.058	4.0472	3.2759	4.8695	60.236	48.197	73.073	18.948	20.153
Washim *	44.631	50.8924	37.9643	38.7811	47.1307	29.891	5.8498	3.7617	8.0731	55.369	49.108	62.036	31.411	32.861
Amravati	42.024	53.8058	29.4562	34.8411	48.1646	20.63	7.1825	5.6413	8.8264	57.976	46.194	70.544	19.109	21.295
Wardha	44.5	54.5566	33.7501	34.8765	47.2043	21.698	9.6238	7.3522	12.052	55.5	45.443	66.25	25.24	26.692
Nagpur	37.819	51.851	22.765	31.5436	46.0357	15.996	6.2756	5.8153	6.7695	62.181	48.149	77.235	14.649	12.947
Bhandara	47.255	52.8448	41.5583	32.167	41.3585	22.8	15.088	11.486	18.758	52.745	47.155	58.442	25.103	25.461
Gondiya *	48.271	53.2737	43.294	32.6383	42.0716	23.253	15.633	11.202	20.041	51.729	46.726	56.706	31.608	32.218
Gadchiroli	51.232	55.4652	46.8943	34.7275	43.6102	25.624	16.505	11.855	21.27	48.768	44.535	53.106	44.115	46.631
Chandrapur	44.942	53.3956	36.0279	32.979	44.3539	20.985	11.963	9.0416	15.043	55.058	46.604	63.972	25.694	25.597
Yavatmal	45.517	53.3249	37.2311	38.6228	48.4289	28.216	6.8944	4.896	9.0153	54.483	46.675	62.769	27.556	29.057
Nanded	41.792	48.7338	34.4193	33.5535	43.3855	23.112	8.2381	5.3483	11.307	58.208	51.266	65.581	32.159	32.358
Hingoli *	46.596	51.3431	41.6167	41.8208	48.0718	35.264	4.7753	3.2712	6.353	53.404	48.657	58.383	44.879	43.794
Parbhani	42.204	49.2706	34.8279	36.6506	46.351	26.526	5.5531	2.9197	8.3015	57.796	50.729	65.172	36.854	35.499
Jalna	44.446	50.2038	38.3885	37.9421	46.4027	29.042	6.5036	3.8012	9.3464	55.554	49.796	61.612	44.328	43.795
Aurangabad	40.623	49.4094	31.1189	35.3996	45.5887	24.378	5.2235	3.8207	6.7409	59.377	50.591	68.881	37.483	33.027
Nashik	43.766	52.7944	34.0306	38.2743	48.8449	26.876	5.4915	3.9495	7.1541	56.234	47.206	65.969	37.687	32.875

Thane	39.105	55.8124	19.6226	34.2112	51.5591	13.981	4.8941	4.2534	5.6412	60.895	44.188	80.377	12.223	8.1137
Mumbai (Suburban) *	36.486	55.953	12.8087	34.3854	53.2377	11.457	2.1002	2.7153	1.3522	63.514	44.047	87.191	0.0946	0.0748
Mumbai	39.297	59.173	13.7228	37.4685	56.8965	12.471	1.8283	2.2765	1.2516	60.703	40.827	86.277	0.1092	0.0831
Raigarh	41.396	52.6958	29.8131	30.1046	43.3304	16.547	11.291	9.3654	13.266	58.604	47.304	70.187	28.554	21.767
Pune	40.85	53.4886	27.0954	36.5767	49.9843	21.986	4.2731	3.5043	5.1097	59.15	46.511	72.905	26.86	20.766
Ahmadna gar	45.959	52.5887	38.9048	39.8714	48.5307	30.657	6.0878	4.058	8.2476	54.041	47.411	61.095	44.379	40.811
Bid	44.307	49.0807	39.2068	38.4341	45.2016	31.204	5.8728	3.879	8.0029	55.693	50.919	60.793	48.348	45.556
Latur	40.404	48.868	31.3485	34.2669	45.7843	21.945	6.1371	3.0837	9.4038	59.596	51.132	68.652	35.687	34.738
Osmanaba d	43.961	50.7172	36.7141	37.3475	46.5954	27.427	6.6138	4.1218	9.287	56.039	49.283	63.286	39.322	41.376
Solapur	45.279	53.0845	36.9295	38.6011	48.7172	27.78	6.678	4.3674	9.1499	54.721	46.915	63.071	34.161	33.436
Satara	46.41	53.669	39.1116	37.2015	48.1218	26.222	9.2086	5.5472	12.89	53.59	46.331	60.888	48.348	43.9
Ratnagiri	44.99	51.1996	39.5229	32.8631	42.5159	24.363	12.127	8.6837	15.16	55.01	48.8	60.477	52.734	39.127
Sindhudur g	46.613	54.6452	39.1693	26.913	38.6439	16.042	19.7	16.001	23.128	53.387	45.355	60.831	46.666	39.896
Kolhapur	46.917	56.5999	36.7166	39.5918	52.1509	26.361	7.3255	4.449	10.356	53.083	43.4	63.283	40.286	35.367
Sangli	47.17	56.0894	37.8512	36.8582	50.3418	22.77	10.312	5.7476	15.081	52.83	43.911	62.149	45.984	43.558

Table 10.16: Demographic Datasets as obtained from the Indian census, 2001 (g)

	CL_PER_F	**AL_PER_P**	**AL_PER_M**	**AL_PER_F**	**HH_PER_P**	**HH_PER_M**	**HH_PER_F**	**OW_PER_P**	**OW_PER_M**	**OW_PER_F**	**CL_MAIN_MARG_P**	**CL_MAIN_MARG_M**	**CL_MAIN_MARG_F**	**AL_MAIN_MARG_P**

Nandurbar *	28.821	47.8535	36.2927	62.8475	1.67942	1.6223	1.7535	16.769	24.628	6.5777	205567	129023	76544	291922
Dhule	25.33	42.804	30.6812	62.2139	2.86039	2.6257	3.2361	26.647	37.531	9.2199	204326	132464	71862	315868
Jalgaon	24.073	46.7414	35.741	65.0489	2.35361	2.2392	2.544	26.8	37.896	8.3338	376999	235682	141317	731028
Buldana	36.218	45.7831	37.4027	57.7042	1.3488	1.3218	1.3872	18.059	27.458	4.6904	353661	201759	151902	465163
Akola	16.478	48.3769	37.7644	70.1457	1.34826	1.1632	1.7279	31.327	40.92	11.649	122831	87824	35007	313600
Washim *	29.341	52.3668	43.1699	65.4934	1.26084	1.3382	1.1504	14.962	22.631	4.0151	143023	87982	55041	238443
Amravati	14.85	51.3099	40.7883	71.81	1.77926	1.6309	2.0683	27.802	36.286	11.272	209362	154179	55183	562163
Wardha	22.731	43.4374	32.1301	62.9766	2.04124	2.0869	1.9624	29.281	39.091	12.33	138909	93052	45857	239058
Nagpur	18.808	24.0002	15.4554	44.8806	2.47558	2.0843	3.4317	58.876	69.514	32.88	225346	141328	84018	369207
Bhandara	24.639	45.3877	35.3367	58.4125	5.36167	3.9814	7.1503	24.148	35.221	9.7982	134773	77155	57618	243680
Gondiya *	30.861	36.3124	29.6134	44.5139	9.82857	5.1285	15.583	22.251	33.04	9.0423	183198	102782	80416	210465
Gadchiroli	41.066	38.1039	28.193	50.1175	1.64694	1.7893	1.4743	16.134	23.387	7.3422	219299	127018	92281	189416
Chandrapur	25.844	39.281	26.8258	58.7453	2.15215	2.3314	1.872	32.873	45.246	13.538	239154	145287	93867	365624
Yavatmal	25.274	49.7088	39.2768	65.566	1.19176	1.2788	1.0594	21.543	30.387	8.1006	308335	196115	112220	556210
Nanded	31.859	41.7109	32.3089	55.8481	2.40849	1.7897	3.339	23.722	33.543	8.9537	386563	233602	152961	501380
Hingoli *	46.283	38.012	30.8705	47.2539	1.46768	1.6379	1.2474	15.641	23.697	5.2159	206434	113635	92799	174847
Parbhani	38.853	37.9681	28.5405	51.8881	1.401	1.3899	1.4173	23.777	34.57	7.8415	237614	136461	101153	244800
Jalna	45.061	33.3801	24.6189	45.433	1.98772	1.7339	2.3369	20.304	29.852	7.1692	317788	181811	135977	239302
Aurangabad	45.137	24.6219	16.9903	37.7294	1.72578	1.383	2.3144	36.169	48.6	14.819	441125	245651	195474	289765
Nashik	45.735	24.7121	17.3188	37.0796	2.18149	1.8138	2.7966	35.42	47.992	14.389	823669	449687	373982	540102
Thane	25.851	9.64292	5.9074	22.0329	2.50404	1.7448	5.0223	75.63	84.234	47.094	388675	198244	190431	306643

Mumbai (Suburban) *	0.1999	0.08612	0.07146	0.16401	3.01493	2.362	6.4841	96.804	97.492	93.152	2982	1984	998	2715
Mumbai	0.2536	0.07402	0.06181	0.14177	3.14819	2.5046	6.7191	96.669	97.35	92.886	1432	924	508	971
Raigarh	40.851	20.3606	13.873	32.1152	2.65844	2.3479	3.221	48.427	62.012	23.813	260980	128193	132787	186095
Pune	39.953	12.6827	7.92077	22.913	2.66541	2.0148	4.0631	57.792	69.298	33.071	793584	418655	374929	374709
Ahmadnagar	49.512	25.3188	18.9533	34.4747	2.9149	2.3795	3.685	27.387	37.857	12.328	824143	447059	377084	470182
Bid	52.083	29.2489	21.9541	39.0054	1.65521	1.6273	1.6926	20.748	30.863	7.219	462976	249608	213368	280083
Latur	37.271	37.4189	28.9411	51.5583	1.79128	1.6249	2.0688	25.102	34.696	9.102	299959	182533	117426	314513
Osmanabad	36.278	39.3445	30.161	52.9531	2.46618	1.9959	3.1631	18.868	26.467	7.6063	256975	161449	95526	257125
Solapur	35.275	28.7415	20.7049	41.0996	4.82553	2.129	8.972	32.272	43.73	14.653	595439	353148	242291	500977
Satara	54.486	21.5557	14.7326	30.9696	2.81324	2.5822	3.132	27.283	38.786	11.412	630296	331808	298488	281013
Ratnagiri	68.254	13.3543	10.3041	16.8337	1.9627	2.1879	1.7058	31.949	48.381	13.206	402561	159161	243400	101945
Sindhudurg	55.42	19.4805	15.8178	24.2159	3.54606	3.2471	3.9326	30.307	41.039	16.432	188992	91105	97887	78893
Kolhapur	48.275	17.7208	12.3286	26.4778	3.57556	3.1633	4.2451	38.418	49.142	21.002	665914	361809	304105	292920
Sangli	49.74	23.3401	17.1158	32.9771	3.32137	2.6412	4.3745	27.354	36.685	12.908	560391	322519	237872	284435

Table 10.17: Demographic Datasets as obtained from the Indian census, 2001 (h)

	AL_MAIN_MARG_M	**AL_MAIN_MARG_F**	**HH_MAIN_MARG_P**	**HH_MAIN_MARG_M**	**HH_MAIN_MARG_F**	**OW_MAIN_MARG_P**	**OW_MAIN_MARG_M**	**OW_MAIN_MARG_F**	**Work Participation Rate**
Nandurbar	125011	166911	10245	5588	4657	102299	84830	17469	46.507

*									
Dhule	139366	176502	21108	11927	9181	196639	170482	26157	43.206
Jalgaon	349173	381855	36810	21876	14934	419146	370224	48922	42.468
Buldana	223147	242016	13704	7886	5818	183487	163815	19672	45.511
Akola	164574	149026	8740	5069	3671	203072	178324	24748	39.764
Washim *	115584	122859	5741	3583	2158	68125	60593	7532	44.631
Amravati	295315	266848	19494	11808	7686	304603	262717	41886	42.024
Wardha	112009	127049	11234	7275	3959	161150	136275	24875	44.5
Nagpur	168715	200492	38083	22753	15330	905713	758830	146883	37.819
Bhandara	107083	136597	28786	12065	16721	129646	106733	22913	47.255
Gondiya *	94473	115992	56966	16361	40605	128967	105405	23562	48.271
Gadchiroli	76795	112621	8187	4874	3313	80202	63703	16499	51.232
Chandrapur	152261	213363	20032	13233	6799	305981	256810	49171	44.942
Yavatmal	265088	291122	13335	8631	4704	241057	205089	35968	45.517
Nanded	233245	268135	28951	12920	16031	285143	242155	42988	41.792
Hingoli *	80101	94746	6751	4250	2501	71946	61488	10458	46.596
Parbhani	109711	135089	9033	5343	3690	153305	132890	20415	42.204
Jalna	102202	137100	14250	7198	7052	145560	123926	21634	44.446
Aurangabad	126372	163393	20310	10287	10023	425657	361481	64176	40.623
Nashik	236896	303206	47678	24810	22868	774124	656464	117660	43.766
Thane	144337	162306	79628	42631	36997	2E+06	2E+06	346920	39.105
Mumbai (Suburban)	1896	819	95046	62666	32380	3E+06	3E+06	465177	36.486

*									
Mumbai	687	284	41296	27836	13460	1E+06	1E+06	186072	39.297
Raigarh	81704	104391	24298	13828	10470	442622	365218	77404	41.396
Pune	159687	215022	78749	40620	38129	2E+06	1E+06	310349	40.85
Ahmadnagar	207624	262558	54131	26066	28065	508590	414701	93889	45.959
Bid	120290	159793	15850	8916	6934	198675	169101	29574	44.307
Latur	152073	162440	15056	8538	6518	210990	182313	28677	40.404
Osmanabad	117689	139436	16117	7788	8329	123305	103276	20029	43.961
Solapur	218681	282296	84111	22486	61625	562514	461867	100647	45.279
Satara	111354	169659	36675	19517	17158	355674	293155	62519	46.41
Ratnagiri	41915	60030	14983	8900	6083	243898	196804	47094	44.99
Sindhudurg	36121	42772	14361	7415	6946	122739	93716	29023	46.613
Kolhapur	126125	166795	59103	32361	26742	635033	502732	132301	46.917
Sangli	126730	157705	40476	19556	20920	333353	271624	61729	47.17